

مركز البحوث

تحليل الانحدار الخطي

تأليف

أ. محمد عبد الرحمن إسماعيل

الطبعة الثانية

منقحة ومزيدة

بسم الله الرحمن الرحيم



مركز البحوث

تحليل الانحدار الخطي

تأليف

أ. محمد عبد الرحمن إسماعيل

الطبعة الثانية
منقحة ومزيدة

١٤٣٧هـ - ٢٠١٦م

بطاقة فهرسة

ⓒ معهد الإدارة العامة، ١٤٣٧هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

إسماعيل، محمد عبدالرحمن

تحليل الانحدار الخطي /محمد عبدالرحمن إسماعيل- الرياض،١٤٣٧هـ

٤٨٨ص؛ ٢١x٢٨ سم.

ردمك: ٩-٢٤٦-١٤-٩٩٦٠

١- الإحصاء الرياضي أ. العنوان

ديوي: ٥١٩,٥ ١٤٣٧/٧٨٠٠

رقم الإيداع: ١٤٣٧/٧٨٠٠

ردمـــــك: ٩-٢٤٦-١٤-٩٩٦٠

المحتـويـات

الموضـــــــــوع		الصفحة

الموضـــــــــوع		الصفحة

الموضـــــــوع		الصفحة

الموضـــــــــوع		الصفحة

مقدمة الطبعة الأولى

إن الهدف الأساسي من معظم البحوث هو تحليل وتقييم العلاقات بين مجموعة من المتغيرات بغرض الوصول إلى صيغة تصف هذه العلاقات. وتعتبر أساليب تحليل الانحدار من أهم وأقوى أساليب التحليل الإحصائي لهذه المتغيرات. وتستخدم أساليب الانحدار في معظم أنواع البحث العلمي، البحث التجريبي وشبه التجريبي والمُشاهد، التي غالباً ما تضم متغيرات تابعة يمكن التنبؤ بها من متغيرات أخرى تعرف بالمتغيرات المفسرة.

يهدف هذا الكتاب إلى تقديم موضوعات **تحليل الانحدار الخطي** من خلال عرض شامل، وسهل، ومتسلسل ومترابط بغرض تنمية مهارات النمذجة الرياضية باستخدام هذا الأسلوب. ولتحقيق هذا الهدف تم تقسيم الكتاب إلى سبعة فصول. ويبدأ الكتاب بعرض بعض المفاهيم الإحصائية المهمة التي تشكل الركيزة الأساسية لموضوعات الفصول اللاحقة. ويتناول الفصل الثاني نموذج الانحدار الخطي البسيط الذي يهدف إلى تحليل العلاقة بين متغيرين أحدهما تابع والآخر مستقل. ويعالج الفصل نموذج الانحدار الخطي المتعدد الذي يستخدم للتنبؤ بقيم المتغير التابع باستخدام متغيرين مستقلين أو أكثر. وتضمن هذا الفصل طريقة تقدير معالم النموذج، الاستدلال الإحصائي، التنبؤ، وفحص النموذج. كما تم التعرض في هذا الفصل إلى موضوعات الانحدار المعياري، الارتباط الجزئي ونماذج الانحدار متعددة الحدود واستخدام بعض حزم برامج الإحصاء الجاهزة في تحليل الانحدار. وفي الفصل الرابع تمت معالجة موضع المشاهدات الشاذة من حيث طرق كشفها وقياس أثرها وبعض طرق معالجتها. أما الفصل الخامس فقد تم تخصيصه لموضوع استخدام المتغيرات النوعية في نموذج الانحدار الخطي حيث تم التعرض إلى طرق ترميز المتغيرات الصورية وبناء نماذج انحدار تضم متغيرات نوعية وكمية ومتغيرات تفاعل بينهما. ويعالج الفصل السادس موضوع اختيار "أفضل" نموذج انحدار عندما يكون لدى الباحث مجموعة كبيرة من المتغيرات المستقلة بهدف الوصول إلى نموذج يضم عدداً قليلاً من هذه المتغيرات ويعطي أعلى درجة من الدقة في التنبؤ بالمتغير التابع. أما الفصل السابع والأخير فيعالج أهم مشاكل الانحراف عن الفروض اللازمة لتقدير معالم نموذج الانحدار الخطي. حيث تناول الفصل مشاكل الارتباط الخطي المتعدد، عدم ثبات تباين حدود الخطأ والارتباط الذاتي بين حدود الخطأ من حيث طرق كشفها والنتائج المترتبة عليها وطرق معالجتها. وقد وضعنا في آخر الكتاب ثلاثة ملاحق هي: التوزيع الطبيعي والتوزيعات المتصلة الأخرى المتصلة بها، والجداول الإحصائية المستخدمة في تحليل الانحدار الخطي، وقائمة بالمصطلحات المستخدمة. كما تم تضمين تمارين في نهاية كل فصل لتمكين القارئ من اختبار مدى فهمه لمادة الفصل.

فبالإضافة للمعالجة النظرية لأساليب الانحدار الخطي تم التركيز في هذا الكتاب على تطبيقاتها، حيث تم في معظم الأمثلة تحليل بيانات حقيقية باستخدام أكثر حزم برامج الإحصاء استخداماً (نظام SAS وSPSS).

وتتطلب قراءة هذا الكتاب إلماماً يسيراً بمبادئ النظرية الإحصائية وحساب التفاضل والتكامل وطرق الجبر الخطي.

وختاماً، إني أرجو من الله أن أكون قد وفقت بهذا الجهد المتواضع إلى توفير مرجع سهل للطالب والباحث. كما أرجو أن يكون هذا الكتاب إضافة حقيقية للمكتبة العربية التي تعاني نقصاً في مثل هذه الكتب. والله الموفق.

مقدمة الطبعة الثانية

يهدف هذا الكتاب في طبعته الثانية إلى تقديم موضوعات تحليل الانحدار الخطي من خلال عرض شامل، وسهل، ومتسلسل ومترابط بغرض تنمية مهارات النمذجة الرياضية باستخدام هذا الأسلوب. ولتحقيق هذا الهدف تمت مراجعة الطبعة الأولى وإضافة العديد من الموضوعات بالإضافة إلى فصل كامل. وتم تقسيم الكتاب إلى ثمانية فصول. ويبدأ الكتاب بعرض بعض المفاهيم الإحصائية المهمة التي تشكل الركيزة الأساسية لموضوعات الفصول اللاحقة. ويتناول الفصل الثاني نموذج الانحدار الخطي البسيط الذي يهدف إلى تحليل العلاقة بين متغيرين، أحدهما تابع والآخر مستقل. ويعالج الفصل الثالث نموذج الانحدار الخطي المتعدد الذي يستخدم للتنبؤ بقيم المتغير التابع باستخدام متغيرين مستقلين أو أكثر. وتضمن هذا الفصل طريقة تقدير معالم النموذج، الاستدلال الإحصائي، التنبؤ وفحص النموذج. كما تم التعرض في هذا الفصل إلى موضوعات الانحدار المعياري، الارتباط الجزئي ونماذج الانحدار متعددة الحدود واستخدام بعض حزم برامج الإحصاء الجاهزة في تحليل الانحدار. وفي الفصل الرابع تمت معالجة موضوع المشاهدات الشاذة من حيث طرق كشفها وقياس أثرها وبعض طرق معالجتها. أما الفصل الخامس فقد تم تخصيصه لموضوع استخدام المتغيرات النوعية في نموذج الانحدار الخطي حيث تم التعرض إلى طرق ترميز المتغيرات الصورية وبناء نماذج انحدار تضم متغيرات نوعية وكمية ومتغيرات تفاعل بينهما. ويعالج الفصل السادس موضوع اختيار "أفضل" نموذج انحدار عندما يكون لدى الباحث مجموعة كبيرة من المتغيرات المستقلة بهدف الوصول إلى نموذج يضم عدداً قليلاً من هذه المتغيرات ويعطي أعلى درجة من الدقة في التنبؤ بالمتغير التابع. وتم تخصيص الفصل السابع لعرض أهم مشكلات عدم استيفاء اشتراطات نموذج الانحدار الخطي وطرق معالجتها. حيث تناول الفصل مشكلات الارتباط الخطي المتعدد، عدم ثبات تباين حدود الخطأ والارتباط الذاتي بين حدود الخطأ من حيث طرق كشفها والنتائج المترتبة عليها وطرق معالجتها. أما الفصل الثامن المضاف في هذه الطبعة فيتناول آخر موضوعين في النمذجة الإحصائية بشكل عام، هما: تأكيد صحة النموذج الذي تم بناؤه وكيفية عرض نتائجه. وفي كل الفصول، ضُمِّنَتْ تمارين بنهاية كل منها لتمكين القارئ من اختبار مدى فهمه لمادة الفصل. وبالإضافة إلى المعالجة النظرية المتعمقة لموضوعات نماذج الانحدار الخطي في الفصول السابقة، تم التركيز كذلك على تطبيقاتها، حيث تم في معظم الأمثلة تحليل بيانات حقيقية وافتراضية باستخدام أكثر حزم برامج الإحصاء استخداماً (نظام SAS وSPSS).

وجاء آخر الكتاب مذيّلاً بثلاثة ملاحق كاشفة ومعينة للقارئ على تحصيل الفائدة العلمية المتكاملة، وهذه الملاحق هي: التوزيع الطبيعي والتوزيعات المتصلة الأخرى المتصلة بها، والجداول الإحصائية المستخدمة في تحليل الانحدار الخطي، وقائمة بالمصطلحات المستخدمة.

وختاماً؛ أتقدم بالشـــكر الجزيل لمقومي الكتاب في الطبعتين الأولى والثانية على الملاحظات العلمية القيمة التي أضافت الكثير إليه، وإلى الزملاء أساتذة الجامعات والطلاب الذين زودوني بملاحظاتهم عن الطبعة الأولى التي تم الأخذ بمعظمها في هذه الطبعة. كما أتقدم بالشـــكر الخاص لسـعادة الدكتور عبدالمحســن اللحيد – مدير عام مركز البحوث السابق- لتشجيعه الدائم لإعداد هذه الطبعة. وأرجو من الله أن أكون قد وفقت بهذا الجهد المتواضع إلى توفير مرجع سـهل للطالب والباحث معاً. كما أرجو أن يكون هذا الكتاب إضـافة حقيقية للمكتبة العربية التي تعاني نقصًا في مثل هذه الكتب.

المؤلف

الرياض ١٤٣٥/٦/٦هـ الموافق ٢٠١٤/٤/٦م

الفصل الأول
مفاهيم أساسية

يتناول هذا الفصل بعض مفاهيم ومصطلحات الإحصاء الأساسية التي ترتكز عليها دراسة تحليل الانحدار الخطي. ويتكون هذا الفصل من أحد عشر موضوعاً، هي: مفهوم الانحدار، استخدامات تحليل الانحدار، مجالات تطبيق نماذج الانحدار، نموذج الانحدار الخطي، المجتمع والعينة، المعلمة وإحصاء العينة، أنواع البيانات، أنواع المتغيرات، مستويات قياس المتغيرات، حجم العينة وعملية النمذجة الرياضية. وقد تبدو هذه الموضوعات منفصلة عن بعضها البعض إلا أنها تشكل الركيزة الأساسية لمادة الكتاب.

١-١ مفهوم الانحدار:

يرجع مصطلح الانحدار (Regression) إلى عالم الأنثروبولوجيا البريطاني السير فرانسيس جالتون (Francis Galton). حيث توصل في ورقة علمية نشرها في أواخر القرن التاسع عشر إلى أنه على الرغم من أن أبناء الآباء الطوال القامة غالباَ ما تكون قاماتهم طويلة وأن أبناء الآباء القصار القامة غالبَا ما تكون قاماتهم قصيرة، إلا أن متوسط أطوال أبناء الآباء طويلي القامة وقصيري القامة يتجه (move) أو ينحدر (Regress) نحو متوسط أطوال أفراد المجتمع ككل (Galton, 1886). ولقد استخدم جالتون مصطلح "الانحدار" للإشارة إلى اتجاه الأطوال نحو المتوسط العام. ويرى جالتون أن الابن يرث بعضًا من صفاته من أبويه ويرث البعض الآخر من أجداده (Ancestry). ولتوضيح العلاقة بين أطوال الأبناء وأطوال آبائهم صاغ جالتون المعادلة التالية (Draper & Smith 1998 p.45):

$$\hat{y} = \bar{y} + \frac{2}{3}(x - \bar{x}) \qquad (1\text{-}1)$$

حيث إن $\hat{Y}$ يمثل تقدير طول الابن و $\bar{Y}$ متوسط أطوال الأبناء و $\bar{X}$ متوسط أطوال الآباء، وX الوسط المرجح لطولي الأب والأم.

وجدير بالذكر أن تحليل جالتون للعلاقة بين أطوال الأبناء وأطوال آبائهم يُعرف اليوم بتحليل الارتباط (Correlation analysis) الذي يهدف إلى قياس قوة العلاقة بين المتغيرات، في حين يستخدم تحليل الانحدار لتحديد صيغة العلاقة بين المتغيرات من خلال إيجاد معادلة تربط بينهما (Draper & Smith, 1998 p.45). وأما تعريف تحليل الانحدار بمفهومه الحديث (انظر Fox 1997 pp16-18 وGujarati and Porter, 2009 p.15) فهو:

"يختص تحليل الانحدار بدراسة اعتماد متغير واحد يعرف بالمتغير المعتمد أو التابع* (Dependent Variable) على متغير واحد أو أكثر تعرف بالمتغيرات المفسرة (Explanatory Variables) أو المتغيرات المستقلة (Independent Variables) أو المنبئات (predictors) وذلك بغرض تقدير و/أو التنبؤ بالقيم المتوسطة للمتغير التابع بمعلومية المتغيرات المفسرة".

---

* يُعرف المتغير التابع أيضًا بمتغير الاستجابة (Response variable).

يستخدم أسلوب الانحدار للتوصل إلى نموذج رياضي يوضح العلاقة بين المتغير التابع، المراد التنبؤ بقيمته، والمتغيرات المفسرة. وجدير بالذكر أن تحليل الانحدار كأسلوب قياس لا يحدد أي المتغيرات يكون تابعًا وأي المتغيرات يكون مستقلاً/مفسراً، وإنما يتم تحديد ذلك بواسطة الباحث مستعيناً في ذلك بالنظريات العلمية وبالدراسات السابقة حول الظاهرة محل البحث وبالملاحظة والخبرة. فالنظرية الاقتصادية مثلًا تشير إلى أن مستوى الادخار يتأثر بسعر الفائدة؛ فكلما زاد سعر الفائدة زاد الادخار، وواضح من هذا المثال أن الادخار متغير تابع وسعر الفائدة متغير مستقل؛ وكذلك الإنفاق يعتمد على الدخل سواء كان ذلك على مستوى الدولة أو الفرد؛ وأن مبيعات سلعة ما تتأثر بمصروفات الدعاية لها والإعلان عنها وبأسعار السلع البديلة وبهامش الربح.. إلخ. كما يمكن أن يُعرف من الملاحظة أن المستوى الأكاديمي للطالب يعتمد على عدد ساعات الاستذكار وعلى مستوى تعليم الأبوين.. إلخ، وأن الأداء الوظيفي للموظف له علاقة بمستوى التعليم وبالخبرة وبطبيعة الوظيفة .. إلخ.

ويلاحظ من تلك الأمثلة أن المتغير التابع يعتمد على المتغيرات المستقلة، أو أن المتغيرات المستقلة تؤثر في المتغير التابع. غير أنه في بعض الحالات نجد أن المتغير التابع يؤثر أيضًا في المتغير المستقل أو المتغيرات المستقلة، أي أن هناك تأثيراً متبادلًا بين المتغيرين التابع والمستقل. فمثلًا الكمية المطلوبة لسلعة ما تتأثر بسعر السلعة ضمن متغيرات مستقلة أخرى، وفي الوقت نفسه لا تخلو الكمية المطلوبة نفسها من تأثير في سعر السلعة.

١-٢ استخدامات تحليل الانحدار:

يستخدم تحليل الانحدار لثلاثة أهداف رئيسة هي:

- **الوصف (Description):** يستخدم نموذج الانحدار لوصف شكل العلاقة بين المتغيرات المفسرة والمتغير التابع، وغالباَ ما يُكتفى في بعض الدراسات الاجتماعية باستخدام نموذج الانحدار لمعرفة المؤشرات فقط؛ وربما يرجع ذلك لمشاكل دقة القياس المرتبطة ببعض المتغيرات الاجتماعية.

- التقدير والتنبؤ (**Estimation and Prediction**): يستخدم نموذج الانحدار لتقدير القيمة المتوسطة والتنبؤ بقيمة مشاهدة جديدة للمتغير التابع المناظرة لقيم فعلية أو متوقعة للمتغيرات المفسرة. ويعتبر التقدير والتنبؤ من أهم استخدامات تحليل الانحدار، إذ تستخدم أساليب الانحدار بشكل مكثف في الواقع العملي لهذا الغرض. فمثلًا ببناء نموذج انحدار خطي لقياس علاقة وزن الطفل (متغير تابع) بعمره (متغير مفسر) لمجتمع أطفال تُراوح أعمارهم من حديثي الولادة إلى سن الخامسة، يمكن للباحث استخدام هذا النموذج لتقدير وزن أي طفل يقع عمره في هذه الفترة المذكورة. كما يستخدم نموذج الانحدار للتنبؤ فقط كما هو الحال في استخدام النموذج للتنبؤ بأحوال الطقس باستخدام بيانات سلاسل زمنية. ولكن يجب ملاحظة أنه كثيراَ ما يساء استخدام نموذج الانحدار بحساب القيم المقدرة للمتغير التابع باستخدام قيم للمتغيرات المستقلة تخرج كثيراَ عن مدى القيم التي استخدمت في بناء النموذج.

- **التحكم (Control):** ويقصد به تفسير التغير في قيم المتغير التابع بدلالة التغير في قيم المتغير المستقل على أساس اتخاذ المتغير المستقل كضابط. والهدف من بناء النموذج هو تحديد الحجم الذي يجب أن يعدل به المتغير المستقل

للحصول على قيمة أو قيم معينة للمتغير التابع. فمثلًا قد يرغب باحث زراعي في تصميم تجربة لمعرفة أثر كمية السماد النيتروجيني (متغير مفسر) على إنتاجية القمح (متغير تابع)، وفي هذه الحالة يحدد الباحث جرعات مختلفة من سماد النيتروجين وتطبيقها على محصول القمح. وببناء نموذج انحدار لتحديد شكل العلاقة بين كمية السماد والإنتاجية؛ يمكن للباحث أن يحدد جرعة السماد التي تحقق الإنتاجية المُثلى.

وفي الواقع العملي نجد أن هذه الاستخدامات الثلاثة متداخلة مع بعضها. فمثلًا يتم بناء نموذج الانحدار بهدف الوصف والتنبؤ معاً أو بهدف الوصف والتنبؤ والتحكم.

١-٣ مجالات تطبيقات نماذج الانحدار:

تهدف معظم البحوث والدراسات إلى تحليل وتقييم العلاقات بين مجموعة من المتغيرات بغرض الوصول إلى صيغة تصف هذه العلاقات. وتُعد طرق تحليل الانحدار من أهم وأقوى أساليب التحليل الإحصائي للمتغيرات. وتستخدم نماذج الانحدار في تطبيقات عديدة في مجالات مختلفة (Frees, 2010; Chatterjee and Hadi, 2013; Bobko, 2001)، منها: علوم الاقتصاد، الإدارة العامة، إدارة الأعمال، التمويل، التسويق، الطب، الأحياء، الكيمياء، الهندسة، الصناعة، الفيزياء، التعليم، الرياضة، الزراعة، علم الاجتماع، علم النفس، العلوم البيطرية، التاريخ، الجغرافيا، علم السكان، والحاسب الآلي، ... وغيرها. ففي مجال الاقتصاد مثلًا يتم بناء نموذج الاستهلاك الأسري كمتغير تابع يمكن تفسيره بمتغيرات مستقلة تشمل: دخل الأسرة، عدد أفراد الأسرة، والحي السكني ... إلخ. وكذلك يستخدم نموذج الانحدار لتحديد المتغيرات المؤثرة في أسعار المنازل والتي تشمل مساحة المنزل، عدد الغرف (نوم، مجالس، ... إلخ)، وموقع المنزل من الشارع، الحي، المسافة بين المنزل ومركز المدينة، عمر المسكن، وغيرها. وفي علم الإدارة يستخدم نموذج الانحدار لمعرفة المتغيرات التي تؤثر في درجة الرضا عن الخدمات الحكومية والتي تضم: مستوى جودة الخدمة، مدة الحصول على الخدمة، سهولة الحصول على الخدمة ونحو ذلك. وفي مجال البحوث الطبية قد يرغب باحث في بناء نموذج انحدار لتحديد المتغيرات المؤثرة في مستوى ضغط الدم لدى البالغين والتي تضم العمر، مستوى الكولسترول، الجنس (ذكر/أنثى)، ممارسة الرياضة (نعم/لا)، مؤشر كتلة الجسم (Body Mass Index)، حالة التدخين (نعم/لا)، وجود أمراض مزمنة أخرى (نعم/لا)، وغيرها. وفي البحوث الزراعية كثيراً ما يتم إجراء تجارب لتحديد كمية السماد (النيتروجين مثلاً) التي تحقق الإنتاجية المثلى للمحصول (القمح مثلاً) باستخدام نماذج الانحدار. وفي الطب البيطري، يمكن بناء نموذج انحدار لقياس أثر كمية الكالسيوم في غذاء الأبقار والفترة بين ولادتين ضمن متغيرات أخرى في إنتاج حليب الأبقار في الموسم. وفي علم السكان تستخدم نماذج الانحدار لمعرفة المتغيرات المؤثرة في مستويات الخصوبة والتي تشمل: عمر الزوج وعمر الزوجة عند الزواج، مستوى تعليم الزوجين، دخل الأسرة، استخدام وسائل تنظيم الأسرة (نعم/لا)، وغيرها من المتغيرات. وفي مجال بحوث التسويق يتم تطبيق نماذج الانحدار لمعرفة أثر حجم المصروفات على الدعاية والإعلان وأسعار السلع البديلة في حجم مبيعات سلعة ما.

١-٤ نموذج الانحدار الخطي (Linear regression model):

تنقسم نماذج الانحدار بصورة عامة إلى **نماذج خطية** و**نماذج غير خطية**. وتنقسم النماذج الخطية إلى نماذج **خطية في المتغيرات** ونماذج **خطية في المعالم** ونماذج **خطية في المتغيرات والمعالم** معاً. وبما أن الاهتمام في هذا الكتاب ينصب على النماذج الخطية، ستناول معناها بشيء من التفصيل.

- **الخطية في المتغيرات المستقلة:**

لقياس العلاقة بين المتغير التابع والمتغيرات المستقلة، تستخدم العلاقة الدالية التالية:

$$Y = f\left(X_1,\ X_2,\ldots,X_p\right) \qquad (1\text{-}2)$$

حيث يشير Y إلى المتغير التابع وf إلى أن المتغير التابع دالة للمتغيرات المفسرة $\left(x_1,\ x_2,\ldots,x_p\right)$. ولكن هذه المعادلة لا تحدد الصيغة الرياضية التي تعبر عن العلاقة التي تربط بين المتغير التابع والمتغيرات المفسرة؛ أي هل هذه العلاقة خطية أم غير خطية؟. ففي حالة الانحدار **خطي المتغيرات** تكون المعادلة الممثلة للعلاقة معادلة من الدرجة الأولى، أي أن يكون أي متغير من متغيرات المعادلة مرفوعًا إلى القوة واحد صحيح وأن لا يكون مضروبًا أو مقسومًا على أي متغير آخر وبذلك يأخذ منحنى الدالة شكل الخط المستقيم عند تمثيلها بيانياً. وبهذا التعريف فإن المعادلة التالية:

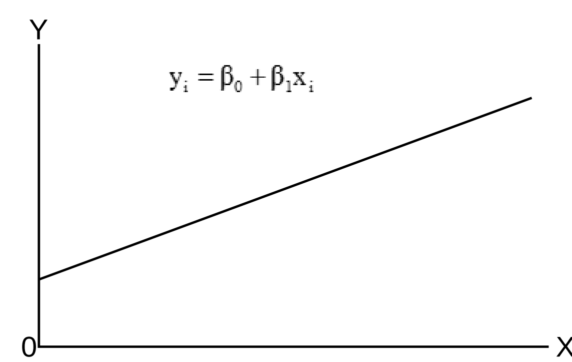
$$y_i = \beta_0 + \beta_1 x_i \qquad (1\text{-}3)$$

حيث إن β_0 وβ_1 معلمتان ثابتتان، تعتبر نموذج انحدار خطي في المتغير لأن أس المتغير المستقل يساوى واحدًا صحيحًا ومنحنى الدالة كما يظهر في الشكل رقم (١-١) خط مستقيم.

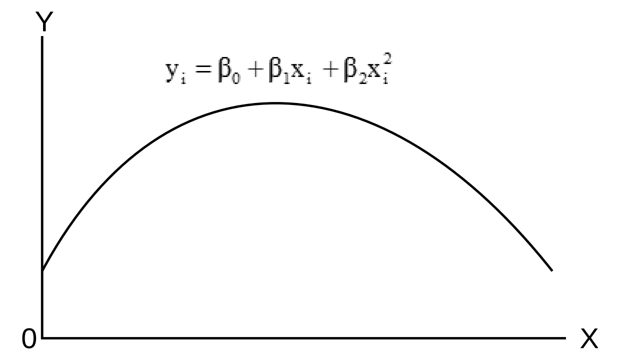
أما الدالة غير الخطية فعبارة عن دالة يكون فيها أحد المتغيرات المستقلة مرفوعاً للقوة غير الواحد الصحيح الموجب أو أن يكون مضروباً أو مقسومًا على متغير آخر أو يظهر كأس، وهي بذلك تأخذ شكل منحنى (خط غير مستقيم) عند تمثيلها بيانياً. ويعتبر نموذج الانحدار التالي:

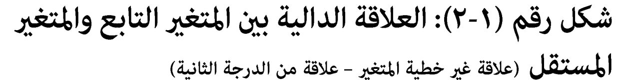
$$y_i = \beta_0 + \beta_1 x_i^2 \qquad (1\text{-}4)$$

نموذج انحدار غير خطي المتغير لأن منحنى الدالة لا يمثل خطًا مستقيمًا؛ ذلك لأن المتغير X مرفوع إلى القوة (٢) (شكل رقم (١-٢)).



شكل رقم (١-١): العلاقة الدالية بين المتغير التابع والمتغير المستقل (علاقة خطية المتغير)




شكل رقم (١-٢): العلاقة الدالية بين المتغير التابع والمتغير المستقل (علاقة غير خطية المتغير - علاقة من الدرجة الثانية)

- **الخطية في المعالم:**

نموذج الانحدار خطي المعالم هو النموذج الذي يكون فيه كل معلمة من معالمه مرفوعة إلى القوة واحد صحيح وغير مضروبة أو مقسومة على أي معلمة أخرى ويمكن أن يكون خطيّا أو غير خطي في المتغيرات. وبدقة أكثر يكون نموذج الانحدار نموذجًا خطي المعالم إذا كان ناتج التفاضل الجزئي لأي معلمة من معالمه لا يتضمن أي معلمة من معالم النموذج $\left(\beta_0, \beta_1, \ldots, \beta_p\right)$. فمثلًا النموذج التالي:

$$y_i = \beta_0 + \frac{1}{\beta_1} x_i \qquad (1\text{-}5)$$

يعتبر نموذجًا غير خطي المعالم لأن المعلمة β_1 مرفوعة إلى القوة سالب واحد صحيح، وكذلك يتضمن ناتج التفاضل الجزئي لـ β_1 المعلمة نفسها $\left(\frac{\partial y_i}{\beta_1} = -\frac{1}{\beta_1^2} x_i\right)$، في حين يعتبر النموذجان (١-٣) و(١-٤) خطيّ المعالم.

- **نموذج الانحدار الخطي:**

نموذج الانحدار الخطي هو نموذج انحدار خطي المعالم، ويمكن أن يكون خطي أو غير خطي المتغيرات. وبهذا التعريف يعتبر النموذجان (١-٣) و(١-٤) نموذجي انحدار خطي في حين يعتبر النموذج (١-٥) نموذج انحدار غير خطي.

١-٥ المجتمع والعينة (Population & Sample):

يختلف معنى كلمة المجتمع في علم الإحصاء عن المعنى الشائع لدى عامة الناس، حيث تستخدم كلمة المجتمع لدى العامة للإشارة إلى جميع الأشخاص الذين يقيمون في منطقة معينة، في حين يُعرّف المجتمع في علم الإحصاء بأنه جميع الوحدات التي تكون الظاهرة محل الدراسة. وبهذا التعريف من الممكن أن يكون المجتمع في علم الإحصاء مجتمعًا بشريًا أو غير ذلك (حيوانات أو جمادات). فمثلًا عندما يرغب باحث في دراسة بعض الجوانب النفسية لمرضى السكر في منطقة

ما، فالمجتمع هنا يتكون من جميع مرضى السكر في تلك المنطقة وكذلك عندما يريد باحث آخر دراسة خاصية معينة لماركة محددة من السيارات، فمجتمع الدراسة هنا يتكون من جميع السيارات من نفس الماركة.

أما العينة فهي جزء من المجتمع يتم اختيارها في الغالب عشوائياَ ومن المفترض أن تُمثل المجتمع محل الدراسة تمثيلًا صادقاً. وهناك اعتبارات عديدة تستدعي دراسة جزء من المجتمع، منها عامل الوقت، التكلفة، تعرض وحدات المجتمع في بعض الحالات للتلف، استحالة دراسة جميع أفراد المجتمع في بعض الحالات.. إلخ (انظر أبو شعر ١٤١٨هـ ص ص ٣٩-٤٥). وتستخدم معظم الدراسات والبحوث الاجتماعية والإنسانية أساليب المعاينة للوصول إلى نتائج يمكن تعميمها على المجتمع.

١-٦ المعلمة وإحصاء العينة (Parameter & Statistic):

المعلمة هي خاصية من خصائص المجتمع التي يتم قياسها كمياَ إذا قمنا بحصر شامل تام ودقيق لكل مفردات المجتمع. فمثلًا نسبة الأشخاص الذين يستخدمون نظارات طبية في المنطقة الشرقية تعتبر معلمة من معالم المجتمع بتلك المنطقة. أما إحصاء العينة (Sample Statistic) فهو قيمة رقمية تصف خاصية معينة يتم قياسها كمياَ عن طريق عينة تمثل مجتمع الدراسة، أي أن إحصاء العينة مُقدر لمعلمة المجتمع. وتستخدم معظم الدراسات والبحوث الاجتماعية أساليب المعاينة لتقدير معالم المجتمع. ففي المثال السابق يمكن للباحث دراسة عينة من أفراد المجتمع بالمنطقة الشرقية لتقدير نسبة الأشخاص الذين يستخدمون نظارات طبية. وفي هذه الحالة تسمى النسبة المُقدرة بإحصاء العينة.

١-٧ أنواع البيانات (Types of Data):

١-٧-١ البيانات المقطعية، بيانات السلاسل الزمنية وبيانات سلسلة زمنية-قطاعية:

يمكن تصنيف البيانات حسب طرق جمعها إلى: بيانات مقطعية (Cross-sectional data)، بيانات سلسلة زمنية (Time-series data)، بيانات سلسلة زمنية-قطاعية/طولية (Panel/Longitudinal data).

- **البيانات المقطعية:** هي البيانات التي يتم جمعها عند نقطة زمنية معينة لمتغير واحد أو عدة متغيرات كالتعداد السكاني الذي يقوم به جهاز الإحصاء المركزي للدولة، المسوحات الاقتصادية الاجتماعية التي يقوم بها الباحثون، المسوحات التي تقوم بها بعض الشركات لمعرف فعالية الإعلانات التي تقوم بها لترويج سلعها... إلخ.
- **بيانات السلاسل الزمنية:** هي البيانات التي تجمع عند فترات زمنية محددة كبيانات أسعار الأسهم اليومية أو الأسبوعية، الحوادث المرورية بالشهور أو السنوات، معدلات هطول الأمطار بالشهور أو السنوات، إحصاءات الأمراض... إلخ.
- **بيانات سلسلة زمنية-قطاعية:** هي البيانات التي تُجمع عبر فترات زمنية متساوية أو مختلفة من نفس الوحدات الإحصائية للمجتمع أو العينة وتعرف أيضًا بالقياسات المتكررة (Repeated measures)، مثال ذلك إخضاع عدد من الأشخاص لعدة اختبارات رياضية لزيادة مهاراتهم في لعبة ما، إذ يتم اختبار أوّليّ لهؤلاء الأشخاص وبعد عدة تمارين يكرر الاختبار مرة ثانية وتستمر التمارين والاختبارات لفترات زمنية متعددة بهدف معرفة أثر التمارين على تنمية مهارات نفس الأشخاص.

ويتم الحصول على البيانات بأنواعها الثلاثة من مصدرين هما: **مصادر أولية** (Primary source) وهي البيانات التي يتم جمعها بشكل خاص لأهداف البحث محل الدراسة وتقوم بنشرها عادة الجهة التي قامت بجمعها وتحليلها، مثال ذلك بيانات التعداد السكاني التي تقوم بجمعها ونشرها الهيئة العامة للإحصاء، **ومصادر ثانوية** (Secondary sources) وهي البيانات الجاهزة المنشورة من قبل لأغراض أخرى وهي مستقاة من مصادر أولية كالنشرات والإحصائية التي تصدرها أجهزة الدولة المختلفة أو تلك التي تنشرها المنظمات الإقليمية والدولية كالمنظمة العربية للتنمية الزراعية، صندوق النقد الدولي، منظمة العمل الدولية وغيرها. وفي أحيانٍ كثيرة يقوم الباحثون باستخدام بيانات ثانوية فقط في إجراء بحوثهم.

١-٧-٢ البيانات المُشاهدة والتجريبية:

تصنف البيانات إلى بيانات مُشاهدة أو غير تجريبية (Observational or non-experimental) وبيانات تجريبية (Experimental data). البيانات المشاهدة هي التي يمكن الحصول عليها عن طريق المشاهدة أو الملاحظة دون التحكم في قيم المتغيرات. وتعتبر المصادر الميدانية الأولية كالمسوحات الاجتماعية الاقتصادية سواء كان ذلك عن طريق المسح الشامل أو باستخدام أسلوب المعاينة والمصادر الثانوية كبيانات السلاسل الزمنية من أهم مصادر الحصول على البيانات المشاهدة. وفي الواقع العملي تستخدم البيانات المشاهدة بكثافة خاصة في مجالات البحوث والدراسات الإنسانية. أما البيانات التجريبية فهي التي يمكن الحصول عليها عن طريق التجربة، أي عن طريق التدبير المُحكم أو السيطرة على واحد أو أكثر من المُتغيرات المـ ستقلة بغرض معرفة تأثيرها على المتغير التابع. ويعتبر التجريب من أهم مصادر البيانات في العلوم الطبيعية. كما يُستخدم التجريب أيضًا في بعض فروع العلوم الإنسانية كعلم النفس والتسويق، ... إلخ. ويلاحظ هنا أن الفرق الأساسي بين البيانات المشاهدة والتجريبية هو أن الباحث في البيانات المشاهدة لا يتعامل مع المتغيرات المستقلة في حين أنه يتحكم فيها في حالة البيانات التجريبية.

١-٨ أنواع المتغيرات (Types of Variables):

تسمى الخصائص التي يشترك فيها جميع أفراد المجتمع الإحصائي وإن اختلفت من وحدة إحصائية إلى أخرى بالمتغيرات. ويسمى المتغير الذي يأخذ قيمًا تحدد بالصدفة وحدها متغيرًا عشوائيًا (Random variable). ويرمز عادة للمتغيرات بالحروف (Y, X, Z, ...). فمثلًا عندما يكون لدينا عدد (n) قيمة تمثل أوزان أطفال حديثي الولادة بمستشفى ما، يمكننا التعبير عن هذه القيم بالمتغير (x_i) حيث يشير الدليل السفلي إلى رقم الوحدة الإحصائية أو المشاهدة، وبالتالي فإن (x_i) تشير إلى المشاهدة رقم i حيث تأخذ أيًا من القيم $(1,2,...,n)$.

والمتغيرات العشوائية نوعان: **متغير عشوائي متقطع** (Discrete random variables) وهو الذي يأخذ قيمًا منفصلة عن بعضها البعض، أي يوجد بينها ثغرات، ومن أمثلة المتغيرات العشوائية المتقطعة: عدد أفراد الأسرة، عدد المرضى المنومين بالشهور في مستشفى ما.. إلخ، مما لا يمكن أن يقاس إلا بالوحدات الكاملة دون الكسور. **والمتغير العشوائي المتصل** (Continuous random variable) هو المتغير الذي يمكن أن يأخذ أي قيمة تقع في نطاق تغيره ولا يتضمن ثغرات كما هو الحال في المتغير المتقطع (انظر الشكل رقم (١-٣)). ومن أمثلة المتغيرات العشوائية المتصلة: الوزن، العمر، الدخل، الاستهلاك، مستوى الكولسترول في دم الإنسان.. إلخ.

كما تصنف المتغيرات إلى متغيرات تابعة وهي التي نحاول تفسيرها والتنبؤ بها، وإلى متغيرات مُفسرة أو مستقلة وهي التي تستخدم في تفسير والتنبؤ بالمتغيرات التابعة.



شكل رقم (١-٣): الاختلاف بين المتغير المتصل والمتقطع

١-٩ مستويات قياس المتغيرات:

تنقسم مستويات قياس المتغيرات إلى أربعة أنواع، هي: المقياس الاسمي (Nominal scale)، المقياس الترتيبي (Ordinal scale)، المقياس الفئوي/الفتري (Interval scale)، مقياس النسب (Ratio scale).

- **المقياس الاسمي:** يستخدم هذا المقياس في حالة الظواهر التي تقاس حسب خاصية معينة مثل الجنس (ذكور – إناث)، الجنسية (سعودي – غير سعودي)، الحالة الزواجية (متزوج - أعزب – أرمل – مطلق). وتستخدم الأرقام أحيانًا لتصنيف الأشياء كأرقام القاعات والمكاتب والسيارات وأرقام الحسابات في البنوك، أرقام لاعبي كرة القدم... إلخ. والأرقام هنا لا تعني أفضلية أو أولوية، وإنما استخدمت فقط لتمييز الخصائص أو الفئات بعضها عن بعض. ولذلك لا يمكن إجراء أي عمليات حسابية كالجمع، الطرح.. إلخ. وبصورة عامة تعتبر كل القياسات الوصفية اسمية بصرف النظر عما إذا كانت الفئات مسماة بأسماء أو أرقام. ويسمى المتغير الاسمي بالمتغير الاسمي الثنائي (Dichotomous variable) إذا كان عدد فئاته يساوي اثنين كمتغير حالة الإصابة بمرض معين (مصاب – غير مصاب) ويسمى بالمتغير الاسمي المتعدد (Polytomous variable) إذا كان عدد فئاته يساوي أكثر من اثنين، مثل متغير الحالة الاجتماعية (متزوج - أعزب – أرمل - مطلق). ولتحليل البيانات يتم عادة ترميز لفئات المتغير الاسمي وذلك بتحويلها إلى أرقام، فمثلًا يتم تحويل متغير الجنس إلى "١" إذا كان الشخص ذكرا، و"٢" إذا الشخص أنثى. وفي تحليل الانحدار يتم تحويل المتغيرات الاسمية إلى متغيرات صورية (Dummy variables) كما سنتعرض لذلك بشيء من التفصيل في الفصل الخامس.

- **المقياس الترتيبي:** يعتبر هذا المقياس أعلى درجة من المقياس الاسمي، إذ لا يستخدم هذا المقياس فقط لتصنيف الأشياء، وإنما ليعكس ترتيبها في تسلسل يبدأ من الأعلى إلى الأسفل أو العكس وفقًا لخصائص معينة يتم قياسها. ويعتبر متغير قياس الرأي "أوافق بشدة"، "أوافق"، "أوافق إلى حد ما"، "لا أوافق"، و"لا أوافق بشدة" من أمثلة القياس الترتيبي الشائعة الاستخدام في البحوث الاجتماعية، حيث يلاحظ أن هناك نوعًا من الترتيب من "أوافق بشدة" إلى

"لا أوافق بشدة". فإذا أعطينا أرقامًا لهذا الترتيب من ٥ إلى ١ على التوالي، فإنه يمكن القول بأن الرقم ٣ أكبر من الرقم ٢ وكذلك الرقم ٤ أكبر من ٣ ولكن يجب ملاحظة أن الفرق بين الرقم ٥ و٤ غير مساو للفرق بين الرقم ٣ و٢ مثلاً، ذلك لأن القياس الترتيبي لا يوضح حجم الفروق بين الرّتب.

- **المقياس الفئوي**: يلي هذا المقياس من حيث الأفضلية المقياس الترتيبي، فبالإضافة إلى الترتيب تكون الفروق بين المستويات المتتالية متساوية تماماً. ففي هذا المقياس يكون الفرق بين الرقم (٧) والرقم (٩) مثلًا مساوياً تمامًا مع الفرق بين الرقم (٢) والرقم (٤). ولكي يكون القياس ذا فئات متساوية لا بد من توافر وحدة قياس معلومة يتم بها قياس قيم المتغير. فوجود خاصية المسافات المتساوية يُمكننا من إجراء العمليات الحسابية كالطرح، القسمة، والوسط الحسابي وخلافه. وأهم ما يميز المقياس الفئوي هو عدم وجود نقطة الصفر المطلق (Absolute zero point)، أي أن الصفر لا يعني غياب الظاهرة أو الخاصية المقاسة. ويعتبر درجات الحرارة المئوية (Celsius) من الأمثلة التقليدية للمقياس الفئوي.
- **مقياس النسب**: يعتبر المقياس النسبي من أعلى مستويات القياس، ويتميز بجميع خصائص المقاييس السابقة، بالإضافة إلى وجود نقطة الصفر المطلق. ووجود نقطة الصفر هذه يمثل الفرق الوحيد بين المقياس النسبي والمقياس الفئوي. والصفر في المقياس النسبي يعبر عن انعدام الظاهرة محل الدراسة. فمثلًا طول قدره صفر يعني أنه لا يوجد طول ولكن درجة حرارة مئوية مقدارها صفر لا يعني عدم وجود حرارة. ومن أمثلة المتغيرات النسبية الوزن، الطول، مستوى الكولسترول في الدم.. إلخ. ولكن يجب ملاحظة أن الفرق بين المقياس الفئوي والنسبي قليل الأهمية في التحليل الإحصائي؛ ذلك لأن كلاً منهما مقاس على مقياس متصل، فضلًا عن أن معظم طرق التحليل الإحصائي لا تميز بين المقياس الفئوي ومقياس النسب.

وتُسمّى المتغيرات التي يتم قياسها اسمياً أو ترتيبياً بالمتغيرات النوعية (Qualitative/Nonmetric variables) والتي يتم قياسها فئوياً أو نسبياً بالمتغيرات الكمية (Quantitative/Metric variables) (شكل رقم (١-٤)). وقد لا يكون من السهل التمييز بين هذه المتغيرات تمييزاً فاصلاً، إذ من المتغيرات ما يمكن التعبير عنه بمقياس ترتيبي وفئوي على حد سواء كما هو الحال، على سبيل المثال، في قياس مستوى ضغط دم الفرد، فيقال مستوى ضغط عالي، متوسط، ضعيف – متغير ترتيبي- أو ذات مستوى محدد (متغير كمي).

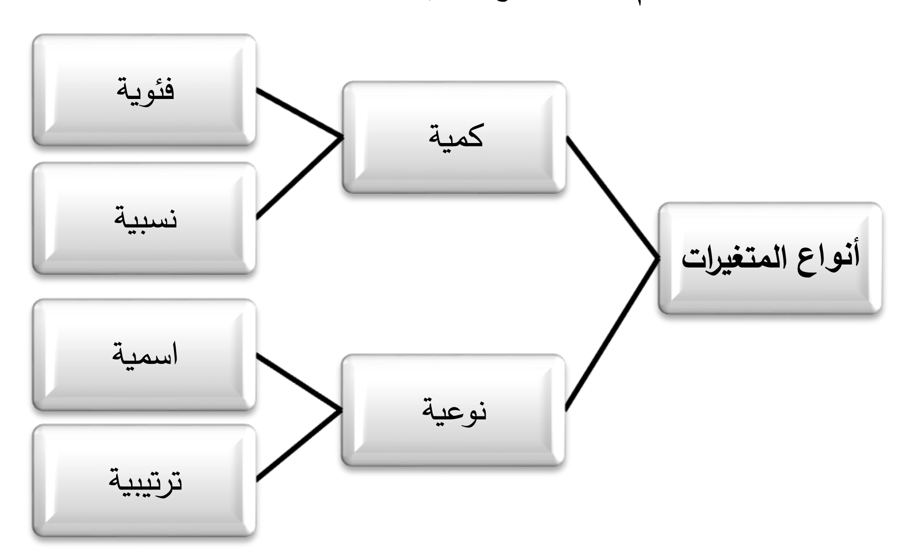
ويجب الإشارة إلى أن طرق التحليل الإحصائي تختلف باختلاف أنواع المتغيرات ومستويات قياسها. ففي تحليل الانحدار الخطي يكون المتغير التابع متغيراً كمياً، ويمكن أن تكون المتغيرات المستقلة كمية أو نوعية، كما يمكن أن تكون نوعية وكمية معاً. ويوضح الجدول رقم (١-١) بعض الطرق الإحصائية لتحليل المتغيرات المتعددة وفقًا لمستويات قياس المتغيرات.

جدول رقم (١-١): بعض النماذج الإحصائية لتحليل المتغيرات المتعددة

النموذج	نوع المتغير التابع	أنواع المتغيرات المستقلة
نموذج الانحدار الخطي المتعدد Multiple linear regression model	كمي (فئوي/ نسبي)	متغيرات كمية (فئوية/ نسبية)، أو كمية ونوعية* معًا
الارتباط الخطي (بيرسون) Pearson Correlation	كمي (فئوي/ نسبي) (لا يفترض تحليل الارتباط وجود متغير تابع وآخر مستقل)	كمي (فئوي/ نسبي)
الارتباط الخطي (سبيرمان) Spearman's Correlation	رتبي / كمي (فئوي/ نسبي) (لا يفترض تحليل الارتباط وجود متغير تابع وآخر مستقل)	رتبي / كمي (فئوي/ نسبي)
تحليل التباين Analysis of variance	كمي (فئوي/ نسبي)	اسمية
تحليل التغاير Analysis of covariance	كمي (فئوي/ نسبي)	خليط من المتغيرات الكمية والاسمية
نموذج الانحدار اللوجستي الثنائي Binary logistic regression	متغير اسمي ثنائي Dichotomous variable	كمية ونوعية
نموذج الانحدار اللوجستي المتعدد Multinomial logistic regression	متغير اسمي متعدد Polytomous variable	كمية ونوعية
التحليل التمييزي Discriminant analysis	اسمي	كمية ونوعية
التحليل العاملي Factor analysis	لا يوجد	يمكن استخدام جميع أنواع المتغيرات. كما يجب ملاحظة انه لا يفترض التحليل العاملي وجود متغير تابع وأخرى مستقلة.
التحليل العنقودي Cluster analysis	لا يوجد	يمكن استخدام جميع أنواع المتغيرات. كما يجب ملاحظة انه لا يفترض التحليل العنقودي وجود متغير تابع وأخرى مستقلة.

* بعد تحويلها إلى متغيرات صورية (Dummy variables)

شكل رقم (١-٤): أنواع المتغيرات ومستويات قياسها



١-١٠ حجم العينة (عدد المشاهدات):

عند بناء نموذج انحدار نعتمد عادة على مشاهدات عينة غالباً ما تكون مسحوبة عشوائياً من مجتمع الدراسة. وللحصول على نتائج يعتد بها ويمكن تعميمها على المجتمع الإحصائي لا بد من اختيار عينة صادقة وممثلة للمجتمع الذي سحبت منه. ولكي تكون العينة ممثلة يتعين على الباحث اختيار حجم عينة مناسب لإجراء أي تحليل إحصائي. ولكن ما هو حجم العينة المناسب؟ توجد عدة طرق للمعاينة التي يمكن تقسيمها إلى المعاينات العشوائية أو الاحتمالية (Probability or Random sampling) والمعاينات غير العشوائية أو غير الاحتمالية (Non-probability or Non-random sampling). وفيما يختص بالمعاينة العشوائية توجد معادلات رياضية مختلفة لتحديد حجم العينة تعتمد على نوع المعاينة، درجة الدقة المطلوبة، مستوى الدلالة والكلفة المالية• . ولكن يلاحظ أن التطور في نظريات المعاينة قد ارتبط بتقدير متغير واحد في حين نجد أن معظم الدراسات والبحوث الاجتماعية تهتم بقياس عدد كبير من المتغيرات. وباختلاف توزيعات هذه المتغيرات واختلاف تباينها فإنه يمكن الحصول على أحجام عينات مختلفة بعدد المتغيرات المراد قياسها. ولذلك من الناحية الإحصائية لا يوجد حجم عينة أمثل بالنسبة للدراسات والبحوث التي تقيس أكثر من متغير واحد. ويقترح كينير وتايلور (كينير وتايلور (١٩٨٣) ص ٣٩٣) في هذا الصدد أن يتم اختيار أكبر حجم للعينة للحصول على درجة دقة أعلى مما هو محدد لها. إلا أن العينات الكبيرة تعاني مشكلة إدارة وجودة البيانات، إذ تزداد أخطاء غير المعاينة مع زيادة حجم العينة هذا فضلًا عن الوقت والتكلفة المالية المضافة لجمع البيانات. ويقترح عدلي علي أبو طاحون (١٩٩٨ ص ٤٤١) أن يتم تحديد حجم العينة وفقاً لخبرة الباحث على أن يكون في حدود ١٠% إلى ٢٠% من حجم المجتمع المراد دراسته وأن لا يقل حجم العينة عن ٣٠

• للمزيد حول موضوع حجم العينة في المعاينات العشوائية يُرجى الرجوع إلى طشطوش (٢٠٠١م) أبو شعر (١٤١٨هـ)، عبدالرحمن أبو عمة، الحسيني عبدالراضي ومحمود هندي (١٤١٥هـ)، Cochran (1977) أو Barnnet (1991) .

مشاهدة. إلا أن هذه النسب قد تأتي أيضًا بحجم عينة كبير إذا كان حجم المجتمع المراد دراسته كبيراَ وربما قد يكون غير ممثلًا لمجتمع الدراسة.

وطبقًا لنظرية المربعات الصغرى – الطريقة التي سيتم استخدامها لتقدير معالم نموذج الانحدار - فإنه يمكن تقدير معالم نموذج الانحدار إذا كان عدد المشاهدات أكبر من عدد المعالم المراد تقديرها. فمثلًا يمكن بناء نموذج انحدار خطي بسيط من ثلاث مشاهدات فقط لأن النموذج يضم معلمتين فقط. إلا أنه لا يتوقع الحصول على نتائج يمكن الاعتماد عليها من عينة حجمها صغير. ويقترح كرستوفر جاتفيلد (Chatfield, 1995; p257) في تحليل الانحدار أن يكون عدد المشاهدات (n) مساوياً على الأقل لأربعة أضعاف عدد المتغيرات المستقلة (p)، أو ما يمكن التعبير عنه بالمتباينة التالية:

$$n \geq 4p$$

ويقترح كل من كلينبان، كوبر ومولر (Kleinbaum, Kupper, and Muller (1988) p318) ونيتر، وزرمان وكتنر (Neter, Wasserman & Kutner (1990) p435) أن يكون عدد المشاهدات على الأقل ما بين خمسة إلى عشرة أضعاف عدد المتغيرات، أي:

$$5p \leq n \leq 10p$$

ويقترح تاباشنك وفيدل (Tabachnick and Fidell, 2007) أن يكون حجم العينة المطلوب لبناء نموذج الانحدار الخطي أكبر من (٥٠) مشاهدة زائدًا ثمانية أضعاف عدد المتغيرات المستقلة (p)، أي:

$$n > (50 + 8 \times p)$$

وعلى الرغم من أنه يمكن الحصول على مقدرات غير متحيزة وكفؤة باستخدام طريقة المربعات الصغرى حتى ولو كان حجمِ العينة (عدد المشاهدات) صغيراً، إلا أن إجراء اختبارات المعنوية الإحصائية يتطلب أن يكون توزيع المتغير التابع طبيعياَ ولذلك كلما كان حجم العينة كبيراَ اقترب توزيع المتغير التابع إلى التوزيع الطبيعي.

١-١١ النمذجة الرياضية (Mathematical Modeling):

تهدف عملية النمذجة إلى ترجمة مشكلة حقيقية إلى وصف رياضي يعرف بالنموذج (Model) بغرض وصفها وتحليلها والتنبؤ بمسارها. ويمكن تقسيم مراحل النمذجة الرياضية إلى سبع مراحل على النحو التالي:

١. **تحديد وصياغة المشكلة:** في هذه المرحلة يتم تعريف المشكلة وحدودها وحجمها؛ وذلك لتكون موضوعًا للبحث والتحليل. وتعد النظرية، الخبرة العلمية والدراسات والبحوث السابقة من أهم مصادر الحصول على المشكلة. ويبدأ الباحث في هذه المرحلة بتلمس أي علاقات قد تربط بين بعض المتغيرات وذلك بطرح عدة تساؤلات. مثلًا: ما العوامل المؤثرة في مستوى ضغط الدم؟ العمر، مؤشر كتلة الجسم (BMI)، الجنس (ذكر/أنثى)، ...، هل توجد علاقة بين مصروفات الدعاية ومبيعات سلعة ما؟ ... ويعتبر تحديد المشكلة والتعرف على طبيعتها وأبعادها شرطًا مسبقًا لأي بحث علمي.

٢. **صياغة فروض لنموذج:** ويقصد بصياغة الفروض تكوين فكرة مبدئية عن النتائج المتوقعة، وذلك بوضع الإجابات المحتملة لأسئلة البحث. فمثلًا عند بناء نموذج انحدار يضع الباحث فروضًا حول العلاقة بين المتغيرات المفسرة والمتغير التابع، وذلك بتكوين فكرة عن الإشارات المتوقعة لمعاملات المتغيرات المفسرة كأن يتوقع الباحث وجود علاقة طردية (موجبة) بين درجة الحرارة ومبيعات الآيس كريم.

٣. صياغة المشكلة رياضياً: تتم في هذه المرحلة صياغة العلاقات بين المتغيرات في صورة رياضية قابلة للقياس، وذلك بتحديد متغيرات النموذج المراد بناؤه ومن ثم تحديد الشكل الجبري للنموذج. وينبع هذا التحديد من النظريات حول الظاهرة محل القياس وخبرة الباحث من الواقع العملي للظاهرة محل الدراسة. وتعد هذه الخطوة من أهم الخطوات، إذ إن الخطأ في تحديد الشكل الجبري للعلاقة بين المتغيرات يترتب عليه أخطاء في قياس وتفسير هذه العلاقة وبالتالي الخروج بنتائج خاطئة.

٤. **حل النموذج الرياضي:** في هذه المرحلة تحدد البيانات المطلوب جمعها وطرق جمعها، اختيار طريقة القياس المناسبة ومن ثم يتم تقدير معالم النموذج.

٥. **تفسير النتائج التي يتم الحصول عليها:** بعد حل النموذج الرياضي يتعين على الباحث تفسير النتائج التي يتم الحصول عليها ومقارنة هذه النتائج بالفروض التي وضعت في المرحلة الثانية ومقارنتها أيضًا بنتائج الدراسات والبحوث السابقة.

٦. **تأكيد صحة النموذج:** عادة ما يتم تأكيد نتائج النموذج بأخذ عينة/عينات أخرى من نفس مجتمع الدراسة وإعادة حل النموذج وذلك للتأكد من مدى استقرار التقديرات بغرض الاطمئنان على ثبات صحة النموذج والنتائج المترتبة عليه.

٧. استخدام النموذج: تستخدم في هذه المرحلة الأخيرة نتائج النموذج الذي تم بناؤه لوصف وتحليل المشكلة موضوع الدراسة والتنبؤ بمسارها بغية الخروج بحلول ومقترحات وتوصيات بشأنها. الشكل رقم (١-٥) يلخص مراحل بناء نموذج رياضي.

شكل رقم (١-٥): مراحل عملية النمذجة الرياضية

المصدر: Burghes & Wood (1980) p.14

تمارين:

١. ما الاختلاف بين النموذج خطي المتغيرات والنموذج خطي المعالم؟ ومن النماذج التالية حدد نماذج الانحدار الخطية مع ذكر السبب؟:

$$y_i = \beta_0 + \beta_1 x_i^{-1} + \varepsilon_i$$
$$y_i = \beta_0 x_i + \beta_1 x_i^2 e^{3x_i} + \varepsilon_i$$
$$y_i = \beta_0 (x_{1i} + \beta_1 x_{2i}) + \varepsilon_i$$
$$y_i = \frac{\beta_0}{1+\beta_1 x_i} + \varepsilon_i$$

٢. إذا أخذت عينة من طلاب البرامج الإعدادية بمعهد الإدارة العامة لمعرفة مستوى الرضا عن البرامج التي يدرسونها، فما هو:

- مجتمع الدراسة.
- معلمة المجتمع وإحصائية العينة المناظرة.

٣. وضح الاختلاف بين مستويات قياس المتغيرات الأربعة (الاسمي، الترتيبي، الفئوي، والنسب). واذكر أنواع قياسات المتغيرات التالية:

- فصيلة الدم (O+, O-, A+,....).
- الحالة الزواجية (متزوج، أعزب، أرمل، مطلق).
- الحالة المرضية (جيدة، وسط، سيئ، سيئ جداً).
- دخل الفرد.
- مدة بقاء مريض بالمستشفى.

الفصل الثاني

نموذج الانحدار الخطي البسيط

٢-١ مقدمـــة:

سنبدأ تحليل الانحدار بدراسة حالة خاصة، حيث يكون لدينا متغيران أحدهما تابع والآخر مستقل والعلاقة بينهما خطية. ويُعرف هذا النوع من الانحدار بنموذج الانحدار الخطي البسيط (Simple linear regression) أو النموذج الخطي لمتغيرين. ويستخدم نموذج الانحدار البسيط للتقدير والتنبؤ بقيم المتغير التابع بدلالة المتغير المستقل. فمثلًا قد يرغب باحث في تحديد شكل العلاقة بين وزن الطفل (متغير تابع) وعمره (متغير مستقل)، العلاقة بين حجم مبيعات سلعة ما (متغير تابع) وحجم مصروفات الدعاية (متغير مستقل)، العلاقة بين مستوى الأداء الوظيفي (متغير تابع) والمؤهل الأكاديمي (متغير مستقل)... إلخ.

وسنشير في نموذج الانحدار الخطي البسيط إلى المتغير التابع بالحرف الإنجليزي (Y) والمتغير المستقل بالحرف (X). ويسمى النموذج في هذه الحالة بنموذج انحدار Y على X. وأن البيانات التي تستخدم في بناء نموذج الانحدار البسيط تحتوي على (n) مشاهدة حول المتغير التابع مع المتغير المستقل (جدول رقم ٢-١).

جدول رقم (٢-١): بيانات نموذج الانحدار الخطي البسيط

رقم المشاهدة	قيم المتغير التابع Y (y_i)	قيم المتغير المستقل X (x_i)
1	y_1	x_1
2	y_2	x_2
3	y_3	x_3
4	y_4	x_4
.	.	.
.	.	.
.	.	.
n	y_n	x_n

٢-٢ شكل الانتشار (Scatter Diagram):

يقدم شكل الانتشار صورة سريعة مرئية لطبيعة العلاقة بين المتغيرين ومدى قوتها واتجاهها، فإننا نستطيع بمجرد النظر إلى الشكل أن نحكم بوجود أو عدم وجود علاقة بين المتغيرين. لذا يُعتبر إعداد الشكل الخطوة الأولى لبناء نموذج الانحدار. ويتم في شكل الانتشار توقيع قيم كل زوج من مشاهدات المتغيرين Y وX [(x_1,y_1)، (x_2,y_2)،..(x_n,y_n)] في شكل نقطة (أو أي علامة أخرى) داخل الفراغ المحصور بين المحور الرأسي والمحور الأفقي. وعادة ما يكون المحور الرأسي لتمثيل المتغير التابع (Y) والمحور الأفقي لتمثيل المتغير المستقل (X)؛ مع مراعاة أن يكون طول المحور الرأسي حوالي ثلاثة أرباع طول المحور الأفقي (Kuiper & Clippinger, 2013). وفي حالة وقوع معظم النقط التي تمثل المشاهدات على خط مستقيم تقريباً، فإننا نقول إن هناك انحداراً خطياً، أي أن هناك علاقة خطية تربط بين المتغيرين. وأما إذا كان الشكل أشبه بمنحنى فنقول إن العلاقة بين المتغيرين غير خطية.

٣-٢ نموذج الانحدار الخطي البسيط:

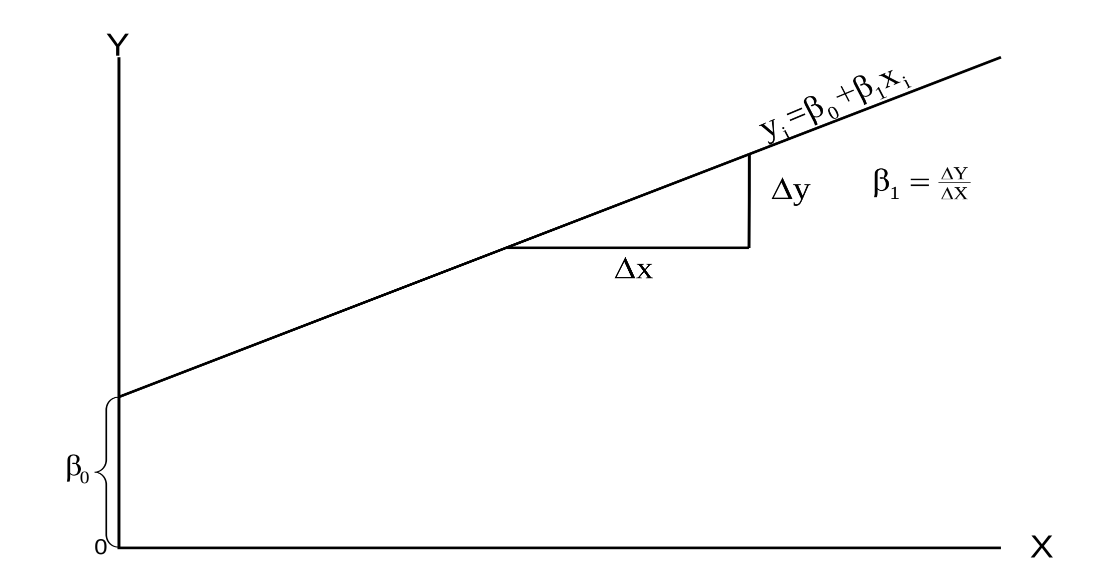
إن أبسط علاقة دالية تربط بين المتغيرين Y وX يمكن التعبير عنها على النحو التالي:

$$Y=f(X) \qquad (2\text{-}1)$$

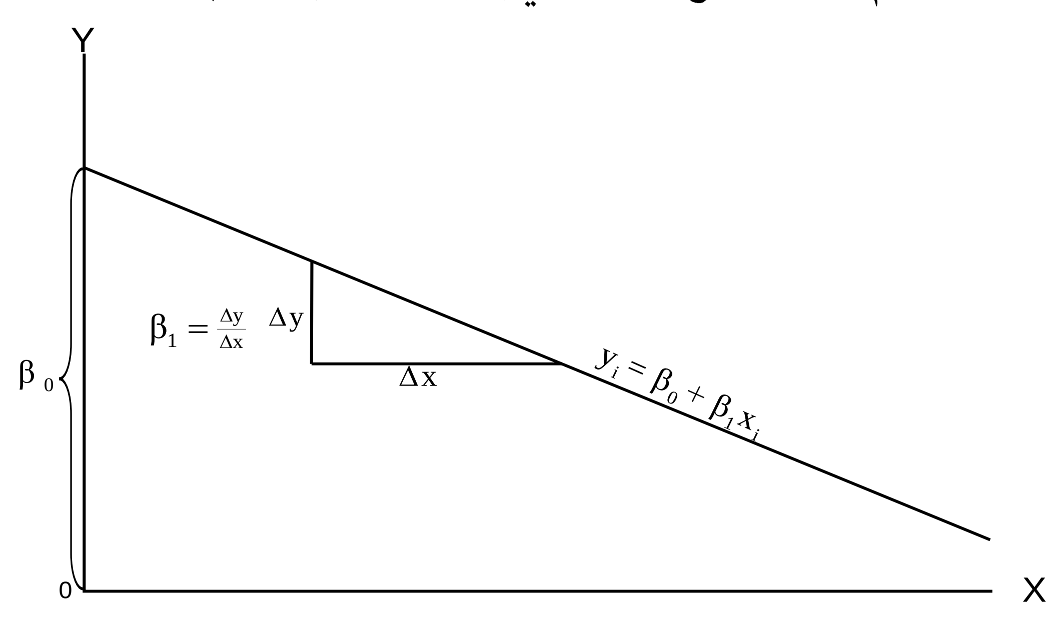
حيث إن f(X) يرمز إلى أن Y دالة (function) لـ X، أي أن قيم Y تتغير تبعًا لتغير قيم X. ولكن هذه المعادلة لا تحدد شكل العلاقة بين المتغيرين، فقد تكون الدالة على صورة خطية أو أي منحنى آخر. ويعتمد تحديد صيغة الدالة على الباحث وذلك بوضع افتراضات حول هذه العلاقة. فمثلًا قد يفترض باحث أن هناك علاقة خطية بين الادخار وسعر الفائدة وبهذا التقريب لشكل الدالة يمكن أن تأخذ المعادلة (2.1) صيغة الدالة الخطية التالية:

$$Y=\beta_0+\beta_1 X \qquad (2\text{-}2)$$

حيث إن Y المتغير التابع وX المتغير المستقل وβ_0 وβ_1 معالم مجهولة تعرف بمعاملات الانحدار. ويعرف β_0 بالمعامل الثابت أو المقطع الصادي (Intercept) وβ_1 بمعامل الانحدار أو ميل الدالة الخطية (Slope) (انظر الشكل رقم (٢-١)). وتسمى المعادلة (2.2) بمعادلة الانحدار الخطي للمجتمع. وبما أن قيمتي المعلمتين β_0 وβ_1 مجهولتان، فإن الهدف هو إذن الحصول على قيم تقديرية لهما على أساس مشاهدات عينة من مجتمع الدراسة.



شكل رقم (٢-ا-أ): نموذج انحدار خطي بسيط - حالة ميل موجب $\beta_1 > 0$



شكل رقم (٢-ا-ب): نموذج انحدار خطي بسيط - حالة ميل سالب $\beta_1 < 0$

ويلاحظ من المعادلة (2.2) أن العلاقة بين المتغير Y وX علاقة تامة ومحددة (شكل رقم (٢-٢-أ)). ولكن في الواقع نجد أن العلاقات بين المتغيرات غالباَ ما تكون غير محددة، كما هو الحال في العلاقات بين المتغيرات العشوائية. فيمكن أن نتصور مثلًا أن هناك علاقة بين دخل الفرد ومصروفاته المعيشية، ولكن لا يمكننا أن نتوقع من أي ريال يكسبه الفرد يصرف منه نسبة محددة (٨٠ هللة مثلًا) على النفقات المعيشية، ومن ثمّ فإن هذه العلاقة بالطبع غير محددة. ولذلك من غير المتوقع عند رسم شكل انتشار المتغيرين Y وX أن تقع جميع النقاط على خط مستقيم، وعلى ذلك تقع بعض النقاط أسفل خط الانحدار ويقع البعض الآخر أعلى هذا الخط (الشكل رقم (٢-٢-ب)). ويمكن تعريف الانحراف عن الخط المستقيم على النحو التالي:

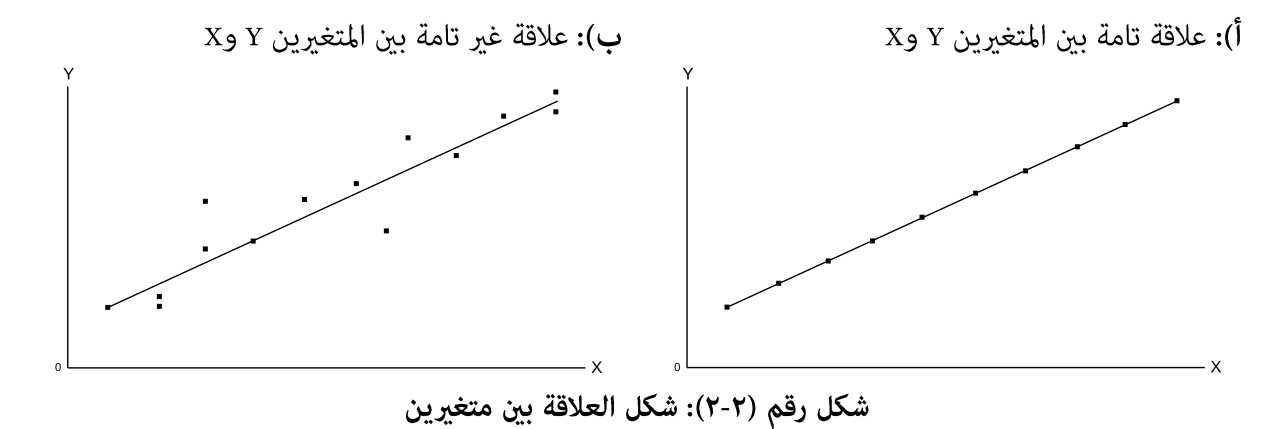
$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i) \qquad (2\text{-}3)$$

حيث إن الانحراف ε_i متغير عشوائي غير مشاهد يأخذ قيمًا سالبة وموجبة ويعرف بحد الخطأ العشوائي (Stochastic error term). وبإضافة حد الخطأ العشوائي يتم تحويل العلاقة من محددة إلى علاقة غير محددة ويأخذ نموذج الانحدار الخطي البسيط الصيغة التالية:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (2\text{-}4)$$

ويمكن إجمال أسباب إدخال حد الخطأ العشوائي في معادلة الانحدار في التالي:

- صعوبة إدخال جميع المتغيرات المستقلة التي تسهم في تفسير تباين المتغير التابع. ويرجع ذلك إلى قلة المعرفة ببعض المتغيرات المؤثرة خاصة في مجالات العلوم الإنسانية والاجتماعية.
- صعوبة قياس بعض المتغيرات المستقلة قياسًا كميَا دقيقًا مثل قياس المتغيرات في الدراسات النفسية.
- عدم وصف نموذج الانحدار بصورة دقيقة.
- وجود أخطاء في قياس المتغيرات (التابع والمتغيرات المستقلة).



شكل رقم (٢-٢): شكل العلاقة بين متغيرين

ويعتمد نموذج الانحدار الخطي البسيط على مجموعة من الاشتراطات نستعرضها فيما يلي:

- عدم وجود أخطاء توصيف للنموذج (No specification error) ويشمل:

  - أن تكون العلاقة بين المتغير التابع والمتغير المستقل علاقة خطية.
  - أن يتضمن نموذج الانحدار الخطي المتغيرات المستقلة التي تسهم في تفسير التباين في المتغير التابع. وهذا يعني عند تحديد النموذج يجب على الباحث أن لا يدخل متغيرات ليس لها تأثير على المتغير التابع وأن لا يجهل في الوقت نفسه إدخال متغيرات ذات تأثير عليه.

- أن تكون قياسات المتغيرين التابع والمستقل صحيحة.

- أن يكون تباين المتغير المستقل X أكبر من الصفر، ويمكن صياغة هذا الفرض كما يلي:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$

حيث إن $\bar{x}$ الوسط الحسابي للمتغير المستقل وn حجم العينة. والهدف من هذه الفرضية هو أن يسهم المتغير المستقل في تفسير التباين في المتغير التابع.

- أن يكون عدد المشاهدات (n) أكبر من عدد معالم نموذج الانحدار المراد تقديرها. ولتقدير معالم نموذج الانحدار الخطي البسيط نحتاج إلى ثلاث مشاهدات أو أكثر.

- الاشتراطات المتعلقة بحد الخطأ العشوائي.

أ) القيمة المتوسطة لحد الخطأ العشوائي تساوي صفرا، أي: $E(\varepsilon_i)=0$. وتعني هذه الفرضية أن يكون مجموع حدود الخطأ السالبة مساوياً لمجموع حدود الخطأ الموجبة بحيث يصبح أثرها في المتوسط يساوي صفراً. فإذا كانت القيمة المتوقعة لحد الخطأ العشوائي لا تساوي صفراً $(E(\varepsilon_i)=d)$ بحيث إن قيمة d لا تساوي الصفر، ستظل مقدرات المربعات الصغرى غير متحيزة لمعاملات الانحدار ما عدا المعامل الثابت (المقطع الصادي) الذي سيكون مقدار تحيزه d وسيصبح المعامل الثابت مساوياً لـ $(\beta_0 + d)$ بدلًا من β_0 وتبعًا لذلك فإن القيمة المتوقعة لـ y_i ستكون أيضًا متحيزة لأن:

$$E(y_i) = (\beta_0 + d) + \beta_1 x_i$$

ب) أن يكون تباين حد الخطأ لكل قيم المتغير المستقل (x_i) ثابتًا. ويمكن صياغة هذه الفرضية رياضيّا كما يلي:

$$\text{var}(\varepsilon_i / x_i) = E\left[\varepsilon_i - E(\varepsilon_i)\right]^2$$

وبما أن القيمة المتوقعة لحد الخطأ تساوي صفراَ $\{E(\varepsilon_i)=0\}$ حسب الفرضية (أ) فإن:

$$\text{var}(\varepsilon_i|x_i) = E(\varepsilon_i^2) = \sigma^2$$

حيث تشير القيمة $var(\varepsilon_i/x_i)$ إلى تباين قيم حد الخطأ المناظرة لقيمة معينة لـ x_i و σ^2 قيمة ثابتة. ويمكن التعبير عن هذه الفرضية لثلاث قيم مختلفة للمتغير المستقل X كما يلي:

$$\sigma^2_{\varepsilon|X_s} = \sigma^2_{\varepsilon|X_r} = \sigma^2_{\varepsilon|X_h}$$

حيث إن s وr وh تشير إلى ثلاث قيم مختلفة للمتغير X، أي أن تباين قيم حد الخطأ العشوائي المناظرة لقيمة المتغير المستقل X_s مساو تمامًا لتباين قيم حد الخطأ العشوائي المناظرة لقيمة المتغير المستقل X_r ولتباين قيم حد الخطأ العشوائي المناظرة لقيمة المتغير المستقل X_h.

ج) استقلال قيم حدود الخطأ بعضها عن بعض. ويتطلب ذلك أن يكون التغاير (covariance) بين حدي الخطأ ε_i و ε_j (i لا تساوي j) يساوي الصفر، أو ما يمكن التعبير عنه رياضيّا بما يلي:

$$cov(\varepsilon_i\varepsilon_j) = E\left[\varepsilon_i - E(\varepsilon_i)\right]\left[\varepsilon_j - E(\varepsilon_j)\right] = 0$$

وبما أن القيمة المتوقعة لحد الخطأ تساوي صفرًا $\left\{E(\varepsilon_i) = E(\varepsilon_j) = 0\right\}$ فإن:

$$cov(\varepsilon_i\varepsilon_j) = E(\varepsilon_i\varepsilon_j) = 0 \qquad \text{for } i \neq j$$

حيث إن i وj مشاهدتان مختلفتان. وبعدم استيفاء هذا الاشتراط تبرز مشكلة تعرف بالارتباط الذاتي بين حدود الخطأ (Autocorrelation). ويرجع سبب هذه الفرضية إلى أننا نود قياس أثر المتغير المستقل فقط على المتغير التابع، فوجود الارتباط الذاتي يعني أن المتغير التابع يعتمد على المتغير المستقل وحد الخطأ معًا. وتظهر هذه المشكلة بشكل متكرر عند تحليل بيانات سلاسل زمنية.

د) استقلالية حد الخطأ (ε_i) عن المتغير المستقل وهذا يعني أن التغاير بين ε_i و x_i يساوي صفرًا، أي:

$$\begin{aligned} cov(x_i\varepsilon_i) &= E\left[x_i - E(x_i)\right]\left[\varepsilon_i - E(\varepsilon_i)\right] \\ &= E\left[x_i\varepsilon_i - x_iE(\varepsilon_i) - E(x_i)\varepsilon_i + E(x_i).E(\varepsilon_i)\right] \end{aligned}$$

وبما أن القيمة المتوقعة لحد الخطأ تساوي صفرًا، فإن:

$$cov(x_i\varepsilon_i) = 0$$

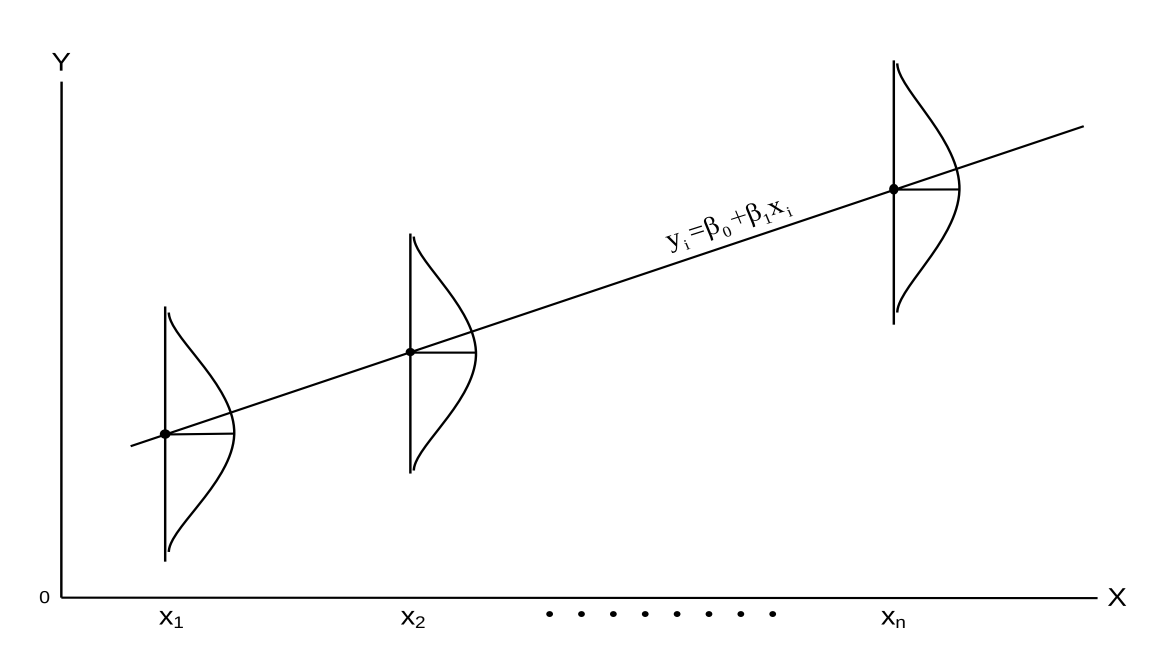
وينص هذا الاشتراط على عدم وجود ارتباط بين حد الخطأ العشوائي والمتغير المستقل وذلك لقياس أثر المتغير المستقل وحده على المتغير التابع. فإذا كان المتغير المستقل يرتبط خطيًا بحد الخطأ العشوائي، فإن ذلك يعني عندما تزيد قيم X_i تزيد قيم ε_i أيضًا. كما أن الارتباط بين X_i و ε_i قد يكون عكسيًا وهذا معناه أن الزيادة في X_i يصاحبها نقص في ε_i. ففي الحالتين توجد صعوبة في فصل تأثير X_i و ε_i على المتغير التابع.

هـ) أن يتبع حد الخطأ العشوائي التوزيع الطبيعي (Normal Distribution). وباستيفاء هذا الاشتراط يمكننا تقييم العلاقة بين المتغير التابع والمتغير المستقل وذلك بإجراء اختبارات الدلالة الإحصائية (فترة الثقة واختبارات الفروض) لمعالم نموذج الانحدار. ويجب أن نشير إلى أنه بالإمكان الحصول على مقدرات "كفؤة" و"غير متحيزة" لمعالم الانحدار حتى في حالة عدم استيفاء هذا الاشتراط. وفي حالة عدم استيفاء هذا الاشتراط، يُنصح بإجراء تحويلة للمتغير التابع كتحويلة اللوغاريتم أو الجذر التربيعي للمتغير (Keinbaum et al 1988 p48).

ويمكن إجمال الاشتراطات الخاصة بحد الخطأ بالصيغة الرياضية التالية:

$$\varepsilon_i \sim NID(0,\ \sigma^2) \qquad (2\text{-}5)$$

أي أن حدود الخطأ مستقلة وتتبع التوزيع الطبيعي بوسط حسابي يساوي صفراَ وتباين ثابت يساوي σ^2 (انظر الشكل رقم (٢-٣).



شكل رقم (٢-٣): توزيع المتغير العشوائي (ε_i) – حالة ثبات التباين

٢-٤ تقدير معالم نموذج الانحدار الخطي البسيط:

يتم عادة استخدام عينة من مجتمع الدراسة لتقدير معالم دالة انحدار المجتمع المجهولة، وتسمى الدالة بدالة انحدار العينة (Sample Regression Function). ويمكن كتابة دالة انحدار العينة المقابلة لدالة انحدار المجتمع (2.4) كما يلي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad (2\text{-}6)$$

حيث إن $\hat{y}_i$ مقدر لـ$E(Y_i|X_i)$ و$\hat{\beta}_0$ و$\hat{\beta}_1$ مقدرتان لـ β_0 و β_1 على التوالي. والآن يمكن التعبير عن دالة انحدار العينة في شكلها العشوائي كما يلي:

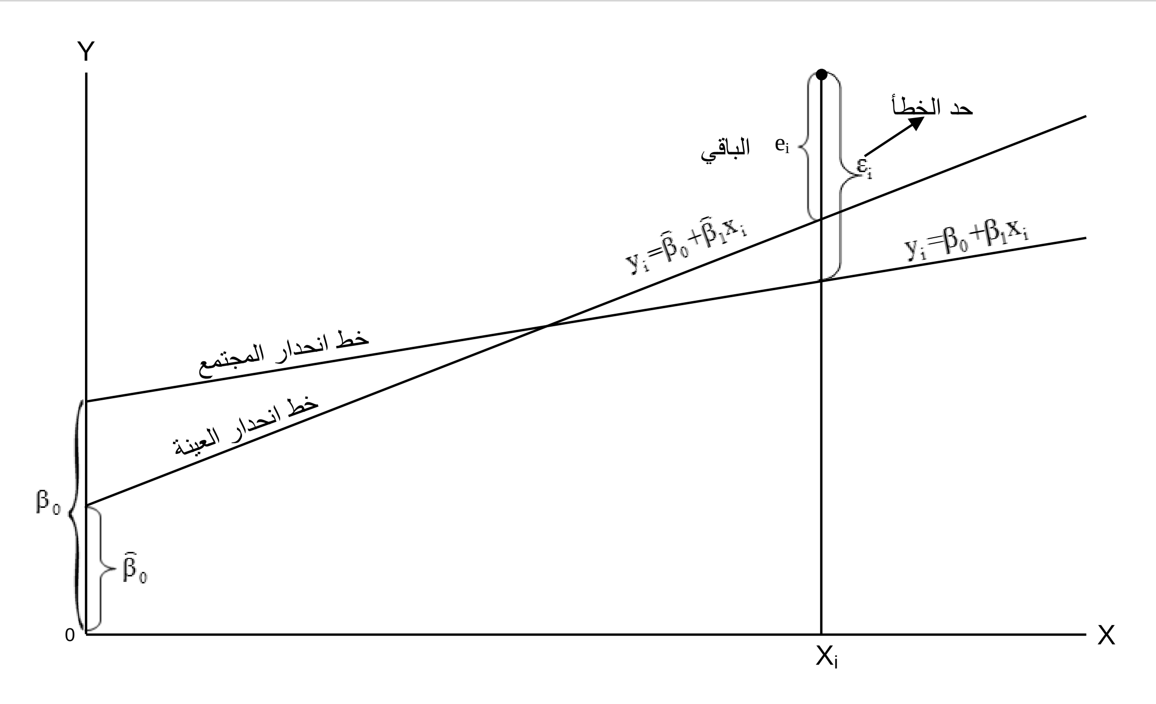
$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad (2\text{-}7)$$

حيث إن y_i القيمة المشاهدة للمتغير التابع و$\hat{\beta}_0$ و$\hat{\beta}_1$ كما جرى تعريفهما أعلاه، ويرمز e_i للباقي، أي الفرق بين القيم المشاهدة أو الفعلية للمتغير التابع والقيمة المقدرة له المبنية على تقديرات معالم نموذج الانحدار من العينة. ورياضيًا يعرف الباقي (e_i) كما يلي:

$$e_i = y_i - \hat{y}_i \qquad (2\text{-}8)$$

حيث إن y_i القيمة المشاهدة الفعلية و $\hat{y}_i$ القيمة المقدرة.

ويرجع إدخال الباقي (e_i) في نموذج انحدار العينة لنفس الأسباب التي ورد ذكرها لإدخال حد الخطأ العشوائي في نموذج انحدار المجتمع. وتجدر الإشارة هنا إلى أن هناك اختلافًا أساسيًا بين الباقي وحد الخطأ. فحد الخطأ هو المسافة الرأسية بين خط انحدار المجتمع غير المشاهد والقيمة المشاهدة للمتغير التابع، في حين يعرف الباقي بأنه المسافة الرأسية بين خط انحدار العينة المقدر والقيمة المشاهدة للمتغير التابع (شكل رقم ٢-٤).



شكل رقم (٢-٤): حد الخطأ والباقي في نموذج الانحدار الخطي البسيط

٢-٤-١ طريقة المربعات الصغرى الاعتيادية (Ordinary Least Squares Method (OLS)):

إذا ما بدا لنا من شكل الانتشار أن العلاقة بين المتغير التابع والمتغير المستقل علاقة خطية تقريباً، فإنه يجب علينا توفيق خط مستقيم واحد يتجمع حوله أكثر النقاط بشكل مقبول. وبالطبع يوجد عدد كبير من الخطوط المستقيمة التي يمكن رسمها ولكن الهدف هو اختيار أفضل خط ملائم للبيانات بحيث تقل انحرافات القيم المشاهدة عن الخط المستقيم إلى أقل حد ممكن. ويمكن الوصول إلى ذلك باستخدام **طريقة المربعات الصغرى الاعتيادية**، الطريقة التي باستخدامها يكون مجموع مربعات انحرافات القيم المشاهدة عن هذا الخط أقل من مجموع مربعات انحرافات القيم عن أي خط مستقيم آخر. وترجع هذه الطريقة إلى عالم الرياضيات الألماني كارل فريدرتش جاوس (Carl Freidrich Gauss). وباستيفاء اشتراطات نموذج الانحدار التي تمت مناقشتها تتميز طريقة المربعات الصغرى بخصائص جيدة جعلتها من أقوى وأوسع الطرق استخداماً. وتعتمد طريقة المربعات الصغرى على تقليل مجموع مربع انحرافات القيم الحقيقية y_i عن القيم المقدرة $\hat{y}_i$ إلى أقل ما يمكن، أي تدنية القيمة التالية:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e^2{}_i \qquad (2\text{-}9)$$

وبما أن $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ ، فأن:

$$\sum_{i=1}^{n} e^2{}_i = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \qquad (2\text{-}10)$$

ويتضح من المعادلة (2.10) أن مجموع مربعات البواقي دالة لمعاملي الانحدار $\hat{\beta}_0$ و $\hat{\beta}_1$، أي:

$$\sum_{i=1}^{n} e_i{}^2 = f(\hat{\beta}_0, \hat{\beta}_1)$$

ولإيجاد قيمتي $\hat{\beta}_0$ و $\hat{\beta}_1$ التي تجعل قيمة مجموع مربعات البواقي إلى أقل حد ممكن، نستخدم التفاضل الجزئي بالنسبة لـ $\hat{\beta}_0$ مرة وبالنسبة لـ $\hat{\beta}_1$ مرة أخرى ومساواة ناتج التفاضل في كل مرة بالصفر كما يلي:

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \qquad (2\text{-}11)$$

و

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \qquad (2\text{-}12)$$

ومن المعادلة (2.11) نحصل على:

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

أي

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \qquad (2\text{-}13)$$

وبإعادة تنظيم المعادلة (2.12) نحصل على:

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \qquad (2\text{-}14)$$

نحصل على التالي:n في (2.14) وضرب المعادلة $\sum_{i=1}^{n} x_i$ في (2.13) وبضرب المعادلة

$$\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i = n\beta_0 \sum_{i=1}^{n} x_i + \beta_1 (\sum_{i=1}^{n} x_i)^2 \qquad (2\text{-}15)$$

و

$$n\sum_{i=1}^{n} x_i y_i = n\beta_0 \sum_{i=1}^{n} X_i + n\beta_1 \sum_{i=1}^{n} x_i^2 \qquad (2\text{-}16)$$

نحصل على:(2.16) من المعادلة (2.15) وبطرح المعادلة

$$n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i = \beta_1 n \sum_{i=1}^{n} x^2{}_i - \left(\beta_1 \sum_{i=1}^{n} x_i\right)^2$$

$$= \beta_1 \left(n\sum_{i=1}^{n} x^2{}_i - \left(\sum_{i=1}^{n} x_i\right)^2 \right)$$

إذن المعامل $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x^2{}_i - (\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (2\text{-}17)$$

وبقسمة طرفي المعادلة (2.13) على n نحصل على:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \qquad (2\text{-}18)$$

إذن المعامل $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (2\text{-}19)$$

حيث إن $\bar{x}$ و $\bar{y}$ الوسط الحسابي للمتغير المستقل X والمتغير التابع Y على التوالي.

ويمكن كتابة معادلة معامل الانحدار $\hat{\beta}_1$ (2.17) بصيغة مجموع المربعات المصحح (corrected sum of squares) كما يلي:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad (2\text{-}20)$$

حيث إن $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$ و $S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} y_i (x_i - \bar{x})$.

ويلاحظ ما يلي على مقدرات المربعات الصغرى:

- تم التوصل إلى هذه المقدرات من مشاهدات عينة مسحوبة من مجتمع الدراسة.
- هذه المقدرات هي مقدرات نقطة (Point Estimators)، بمعنى أنها تعطي تقديراً واحدًا لكل معلمة من معالم النموذج. وسنتعرض في الجزء (٢-٩-٢) للتقدير بفترة (Interval Estimation) لهذه المقدرات.

٢-٤-٢ طريقة الإمكان الأعظم (Maximum Likelihood Estimation (MLE)):

يمكن الحصول على نفس مقدرات المربعات الصغرى باستخدام طريقة الإمكان الأعظم الطريقة الأكثر استخداماً في تقدير معالم نماذج الانحدار غير الخطية. وتتطلب طريقة الإمكان الأعظم أن يكون توزيع الأخطاء طبيعياً على عكس طريقة المربعات الصغرى التي لا تتطلب شكلًا محدداً لتوزيع الأخطاء(ε) لتقدير معالم نماذج الانحدار الخطي. وتأخذ دالة الإمكان في نموذج الانحدار الخطي البسيط لمجموعة البيانات $(y_i , x_i ; i=1,2,..n)$ حيث الصيغة التالية (Montgomery, Peck and Vining 2012; Yan and Su, 2009):

$$\begin{aligned} L(y_i, x_i, \beta_0, \beta_1, \sigma) &= \prod_{i=1}^{n}\left(2\pi\sigma^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2} \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2} \end{aligned} \qquad (2\text{-}21)$$

ومقدرات طريقة الإمكان الأعظم لمعالم نموذج الانحدار الخطي البسيط $\left(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2\right)$ يمكن الحصول عليها بتعظيم الدالة L بأخذ لوغاريتم الدالة (2.21) كما يلي:

$$\ln L(y_i, x_i, \beta_0, \beta_1, \sigma) = -\left(\tfrac{n}{2}\right)\ln 2\pi - \left(\tfrac{n}{2}\right)\ln\sigma^2 - \left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \qquad (2\text{-}22)$$

وباستخدام التفاضل الجزئي بالنسبة لـ β_0 مرة وبالنسبة لـ β_1، ولـ σ^2 مرة أخرى ومساواة ناتج التفاضل في كل مرة بالصفر نحصل على مقدرات الإمكان الأعظم كما يلي:

$$\left.\frac{\partial \ln L}{\partial \beta_0}\right|_{\tilde{\beta}_0,\tilde{\beta}_1,\tilde{\sigma}^2} = \frac{1}{\tilde{\sigma}^2}\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right) = 0$$

و

$$\left.\frac{\partial \ln L}{\partial \beta_1}\right|_{\tilde{\beta}_0,\tilde{\beta}_1,\tilde{\sigma}^2} = \frac{1}{\tilde{\sigma}^2}\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right)x_i = 0$$

و

$$\left.\frac{\partial \ln L}{\partial \sigma^2}\right|_{\tilde{\beta}_0,\tilde{\beta}_1,\tilde{\sigma}^2} = -\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right)^2 = 0$$

وبحل المعادلات الثلاث نحصل على مقدرات الإمكان الأعظم وهي مساوية تمامًا لمقدرات المربعات الصغرى، وهي كما يلي:

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x} \qquad (2.23)$$

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (2.24)$$

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i\right)^2}{n} \qquad (2.25)$$

ويلاحظ أن تقدير معلمتي نموذج الانحدار الخطي البسيط باستخدام طريقة الإمكان الأعظم بافتراض تبعية الأخطاء للتوزيع الطبيعي مساوٍ لتقدير المعلمتين باستخدام طريقة المربعات الصغرى، أي أن $\hat{\beta}_0 = \tilde{\beta}_0$ و $\hat{\beta}_1 = \tilde{\beta}_1$. أما مقدر التباين باستخدام طريقة الإمكان الأعظم متحيز لـ σ^2، ولكن يلاحظ أن حجم التحيز يقل كلما كان حجم العينة كبيراً. وبصورة عامة تتميز مقدرات الإمكان الأعظم بخصائص إحصائية أفضل من مقدرات طريقة المربعات الصغرى (Montgomery et al., 2012). فمقدرات الإمكان الأعظم غير متحيزة عدا مقدر التباين الذي يقترب من عدم التحيز في حال كبر حجم العينة ولديها أقل تباين مقارنة بالمقدرات الأخرى، وتتسم بالاتساق (Consistent estimators)، وبالكفاية (Sufficient). كما أن طريقة الإمكان الأعظم تستخدم لتقدير معالم النماذج غير الخطية (Non-linear models).

٢-٥ تفسير معاملات نموذج الانحدار الخطي البسيط:

يمثل المعامل الثابت $\hat{\beta}_0$ القيمة المتنبأ بها للمتغير التابع (Y) عندما تكون قيمة المتغير المستقل (X) مساوية للصفر $(x_i = 0)$، أي رياضياً تصبح القيمة المقدرة لـ y_i كما يلي:

$$\hat{y}_i = \hat{\beta}_0$$

ولكن عند تفسير المعامل الثابت يجب مراعاة ما يلي:

- من الخطأ أن نقول عندما تكون قيمة المتغير المستقل (X) مساوية للصفر تكون قيمة المتغير (Y) مساوية للمعامل الثابت إذا كانت قيم المتغير المستقل التي استخدمت في بناء النموذج أكبر من الصفر. ويجب أن يكون التنبؤ بقيم المتغير التابع (Y) بالتعويض في نطاق قيم مشاهدات المتغير المستقل.
- المشكلة الأخرى في تفسير المعامل الثابت هي عندما تكون إشارة المعامل الثابت سالبة وقيم المتغير التابع موجبة أو العكس عندما تكون إشارة المعامل الثابت موجبة وقيم المتغير التابع سالبة. وترجع صعوبة التفسير في الحالة الأولى إلى أنه إذا كانت قيمة المتغير المستقل مساوية للصفر فإن القيمة المقدرة للمتغير التابع (Y) ستكون سالبة في حين نجد أن قيم المتغير التابع موجبة كالدخل، العمر الوزن وغيرها. وفي هذه الحالة لا معنى لتفسير المعامل الثابت؛ ويستخدم المعامل الثابت $\hat{\beta}_0$ مع المعامل $\hat{\beta}_1$ للتنبؤ بقيم المتغير التابع.

أما المعامل $\hat{\beta}_1$ الذي يعرف بميل الانحدار ويمثل مقدار التغير الذي نتنبأ به بالنسبة للمتغير التابع لكل تغير مقداره وحدة واحدة في المتغير المستقل X. فإذا كانت إشارة المعامل $\hat{\beta}_1$ موجبة فإن ذلك يعني أن العلاقة التي تربط بين المتغيرين علاقة طردية، أي أن أي زيادة في قيمة المتغير المستقل بوحدة واحدة تؤدي إلى زيادة في قيمة المتغير التابع بمقدار ثابت يساوي $\hat{\beta}_1$. وأما إذا كانت إشارة المعامل $\hat{\beta}_1$ سالبة فنقول إن العلاقة بين المتغيرين علاقة عكسية، أي أن أي زيادة في قيمة المتغير المستقل بوحدة واحدة تؤدي إلى انخفاض في قيمة المتغير التابع بمقدار $\hat{\beta}_1$.

٢-٦ خصائص خط الانحدار المقدر:

- متوسط القيمة المقدرة للمتغير التابع مساو لمتوسط قيم المتغير التابع الفعلية، أي أن:

$$\bar{\hat{y}} = \bar{y}$$

البرهان:

من المعادلة (2.6) نعلم أن:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

وبجمع وقسمة طرفي المعادلة على n نحصل على:

$$\frac{\sum_{i=1}^{n} y_i}{n} = \frac{n\hat{\beta}_0}{n} + \frac{\hat{\beta}_1 \sum_{i=1}^{n} x_i}{n} = \bar{\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

و بما أن $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ حسب المعادلة (2.18) فإن:

$$\bar{\hat{y}} = \bar{y}$$

- متوسط البواقي يساوي صفراَ، أي أن: $\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} = 0$

البرهان:

بما أن:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i = n\bar{y} - n\bar{\hat{y}}$$

ومن الخاصية الأولى أثبتنا أن $\bar{\hat{y}} = \bar{y}$ فإذن:

$$\sum_{i=1}^{n} e_i = n\bar{y} - n\bar{y} = 0$$

وبالتالي فإن متوسط البواقي يساوي صفراَ.

- يمر خط الانحدار بنقطة متوسطي العينة للمتغيرين التابع والمستقل.

البرهان:

من الخاصية الأولى والمعادلة (2.18) نجد أن:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

- البواقي غير مقترنة بالقيمة المقدرة للمتغير التابع. ويعني ذلك أن التغاير بـين البـواقي والقـيم المقـدرة للمتغـير التابع يساوي صفراَ، أي:

$$\operatorname{cov}(\hat{y}_i e_i) = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})(e_i - \bar{e}) = 0$$
$$= \sum_{i=1}^{n} \hat{y}_i e_i = 0$$

البرهان:

للبرهنة على أن التغاير بين البواقي والقيمة المقدرة للمتغير التابع يساوي صفراَ، لا بد من بعض الإثباتات الأولية التالية:

بطرح المعادلة (2.18) من المعادلة (2.6) نحصل على:

$$\hat{y}_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x})$$

وبالتالي نحصل على:

$$\hat{y}_i = \hat{\beta}_1 (x_i - \bar{x}) + \bar{y}$$

والآن يمكن إثبات أن التغاير بين البواقي وقيمة المتغير التابع المقدر يساوي صفراَ كما يلي:

$$\sum_{i=1}^{n} \hat{y}_i e_i = \sum_{i=1}^{n} \left(\hat{\beta}_1 \left(x_i - \bar{x} \right) + \bar{y} \right) e_i$$
$$= \hat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \bar{x} \right) e_i + \bar{y} \sum_{i=1}^{n} e_i$$

وبما أن $\bar{y} \sum_{i=1}^{n} e_i = 0$ فإن

$$\sum_{i=1}^{n} \hat{y}_i e_i = \hat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \bar{x} \right) e_i$$

ومن ثم فإن

$$\sum_{i=1}^{n} \hat{y}_i e_i = \hat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \bar{x} \right) \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)$$
$$= \hat{\beta}_1 \left\{ \sum_{i=1}^{n} x_i y_i - \bar{x} \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \bar{x} \hat{\beta}_1 \sum_{i=1}^{n} x_i - n \bar{x} \hat{\beta}_0 + n \bar{x} \hat{\beta}_0 \right\}$$
$$= \hat{\beta}_1 S_{xy} - \hat{\beta}_1^2 S_{xx}$$

وبما أن $S_{xy} = \hat{\beta}_1 S_{xx}$ فإن

$$\sum_{i=1}^{n} \hat{y}_i e_i = \hat{\beta}_1 S_{xy} - \hat{\beta}_1^2 S_{xx} = \hat{\beta}_1^2 S_{xx} - \hat{\beta}_1^2 S_{xx} = 0$$

- البواقي غير مقترنة بالمتغير المستقل، ويعني هذا أن التغاير بين المتغير المستقل والبواقي يساوي صفراَ، أي:

$$cov(x_i e_i)=\sum_{i=1}^{n}(x_i-\bar{x})(e_i-\bar{e})=\sum_{i=1}^{n}x_i e_i=0$$

يمكن كتابة التغاير بين المتغير المستقل والبواقي كما يلي:

$$\sum_{i=1}^{n}x_i e_i=\sum_{i=1}^{n}x_i\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)$$

$$=\sum_{i=1}^{n}x_i y_i-\hat{\beta}_0\sum_{i=1}^{n}x_i-\hat{\beta}_1\sum_{i=1}^{n}x_i^2$$

وبما أن $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$ و $S_{xy}=\hat{\beta}_1 S_{xx}$ فإن:

$$\sum_{i=1}^{n}x_i e_i=\sum_{i=1}^{n}x_i y_i-\left(\bar{y}-\hat{\beta}_1\bar{x}\right)\sum_{i=1}^{n}x_i-\hat{\beta}_1\sum_{i=1}^{n}x_i^2$$

$$=\sum_{i=1}^{n}x_i y_i-n\bar{y}\bar{x}-\hat{\beta}_1\sum_{i=1}^{n}x_i^2+\hat{\beta}_1 n\bar{x}^2$$

$$=S_{xy}-\hat{\beta}_1 S_{xx}=\hat{\beta}_1 S_{xx}-\hat{\beta}_1 S_{xx}=0$$

مثال:

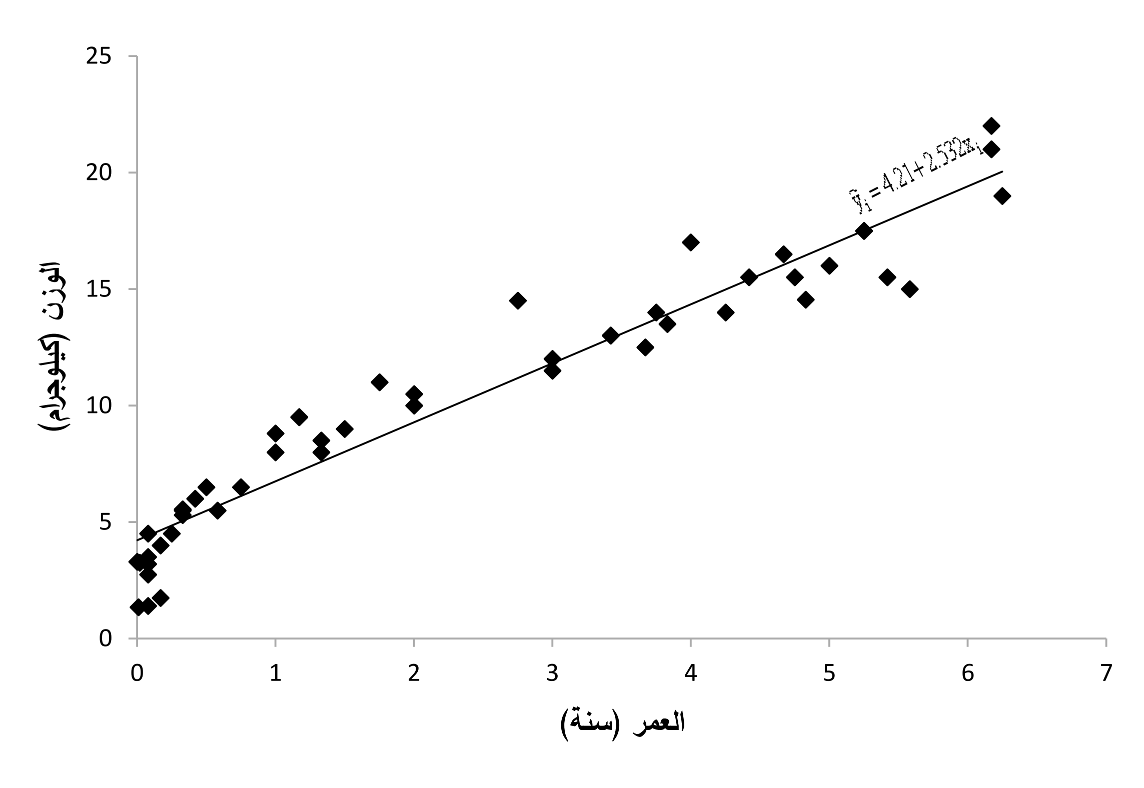
يوضح الجدول رقم (٢-٢) بيانات عن أوزان (٥٠) طفلًا تُراوح أعمارهم بين يوم واحد وست سنوات وثلاثة شهور أخذت بصورة عشوائية من سجلات مستشفى أبها للنساء والولادة والمراكز الصحية التابعة لوزارة الصحة بالمملكة العربية السعودية. وفي هذا المثال سنقوم ببناء نموذج انحدار خطي بسيط لقياس العلاقة بين وزن الطفل (متغير تابع) وعمره (متغير مستقل) بهدف تقدير أو التنبؤ بوزن الطفل في هذه الفترة من عمره.

الحل:

لبناء نموذج انحدار وزن الطفل على عمره نبدأ أولًا برسم شكل انتشار المتغيرين كما هو موضح بالشكل رقم (٢-٥). حيث يلاحظ من الشكل أن هناك علاقة خطية واضحة بين المتغيرين. ولذلك يعتبر نموذج الانحدار الخطي البسيط نموذجًا ملائمًا لتوفيق البيانات. ويأخذ النموذج الصيغة التالية:

$$y_i=\beta_0+\beta_1 x_i+e_i$$

حيث y_i وزن الطفل (كيلوجرام) و x_i عمر الطفل (سنة) و e_i الباقي.



شكل رقم (٢-٥): شكل انتشار وزن الطفل مع عمره

ويوضح الجدول رقم (٢-٣) الحسابات المطلوبة لتقدير معالم نموذج الانحدار. ومن الجدول تم حساب القيم التالية:

$$\sum_{i=1}^{50} x_i y_i = 1739.226 , \quad \sum_{i=1}^{50} x^2{}_i = 491.6858 , \; S_{xx} = 216.218408, \; \sum_{i=1}^{50} x_i = 117.360$$

$$\sum_{i=1}^{50} y_i = 507.70 , \; S_{xy} = 547.55256 , \; \bar{y} = 10.154, \; \bar{x} = 2.3472$$

والآن لإيجاد قيمتي معاملي الانحدار $\hat{\beta}_0$ و $\hat{\beta}_1$ يتم التعويض في المعادلتين (2.17) و(2.19) على النحو التالي:

أولًا: المعامل $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{50} x_i y_i - \sum_{i=1}^{50} x_i \sum_{i=1}^{50} y_i}{n\sum_{i=1}^{50} x^2{}_i - \left(\sum_{i=1}^{50} x_i\right)^2} = \frac{(50 \times 1739.226 - 117.360 \times 507.70)}{(50 \times 491.6858 - 117.360 \times 117.360)} = 2.532$$

كما يمكن الحصول على نفس التقدير باستخدام المعادلة (2.20) كما يلي:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{547.55256}{216.218408} \approx 2.532$$

ثانيًا: المعامل الثابت $\hat{\beta}_0$:

وبالتعويض في الصيغة (2.19) نحصل على قيمة المعامل الثابت:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 10.154 - 2.532x2.3472 = 4.21$$

وعلى ذلك فإن نموذج انحدار وزن الطفل على عمره المقدر هو:

$$\hat{y}_i = 4.21 + 2.532x_i$$

ويمكن تفسير النموذج المقدر كما يلي:

- المعامل الثابت $\hat{\beta}_0$ يمثل القيمة المتوقعة للمتغير التابع عندما تكون قيمة المتغير المستقل مساوية للصفر. وبما أن عينة الدراسة تحتوي على أطفال حديثي الولادة، أي أن قيم المتغير المستقل (العمر) تحتوي على قيم قريبة جدًا من الصفر، فيمكننا تفسير المعامل الثابت على أنه متوسط وزن الطفل عند الولادة والذي يقدر بـ ٤,٢١ كيلوجرام في عينة الدراسة بمدينة أبها.
- المعامل أو ميل الانحدار المقدر $\hat{\beta}_1$ يقيس التغير في وزن الطفل الناتج عن تغير العمر بوحدة واحدة (سنة). وبمعنى آخر تُقدر الزيادة السنوية في وزن الطفل بـ ٢,٥٣٢ كيلوجرام في الفترة من تأريخ الولادة إلى أن يصل الطفل سن السادسة تقريبًا. ويرسم عادة خط الانحدار الموفق على شكل الانتشار لتوضيح مدى توفيق النموذج للبيانات المشاهدة كما يوضح ذلك الشكل رقم (٢-٥).

جدول رقم (٢-٢): أوزان وأعمار (٥٠) طفلًا من منطقة عسير

رقم المشاهدة	وزن الطفل (كيلوجرام)	عمر الطفل (سنة)
1	11.50	3.00
2	16.00	5.00
3	6.50	0.50
4	17.00	4.00
5	8.50	1.33
6	8.80	1.00
7	22.00	6.17
8	13.00	3.42
9	12.50	3.67
10	15.50	5.42
11	9.50	1.17
12	15.50	4.42
13	9.50	1.17
14	14.50	2.75
15	19.00	6.25
16	9.00	1.50
17	14.00	4.25
18	10.50	2.00
19	6.00	0.42
20	15.00	5.58
21	13.00	3.42
22	21.00	6.17
23	12.00	3.00
24	17.50	5.25
25	5.50	0.33
26	5.30	0.33
27	6.50	0.75
28	13.50	3.83
29	4.50	0.25
30	15.50	4.75
31	16.50	4.67
32	11.00	1.75
33	17.50	5.25
34	14.55	4.83
35	10.00	2.00
36	4.00	0.17
37	3.50	0.08
38	8.00	1.00
39	8.00	1.33
40	14.00	3.75
41	1.75	0.17
42	3.20	0.08
43	5.55	0.33
44	2.75	0.08
45	1.35	0.01
46	5.50	0.58
47	4.50	0.08
48	3.25	0.02
49	3.30	0.00
50	1.40	0.08

المصدر: مستشفى أبها للنساء والولادة والمراكز الصحية، وزارة الصحة، المملكة العربية السعودية (١٤٢٢هـ)

جدول رقم (٢-٣): الحسابات اللازمة لتقدير معالم نموذج وزن الطفل على عمره

المشاهدة	y_i	x_i	$x_i y_i$	x_i^2	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(y_i-\bar{y})(x_i-\bar{x})$	$(x_i-\bar{x})^2$
1	11.50	3.00	34.5000	9.0000	0.65280	1.34600	1.81172	0.87867	0.42615
2	16.00	5.00	80.0000	25.0000	2.65280	5.84600	34.17572	15.50827	7.03735
3	6.50	0.50	3.2500	0.2500	-1.84720	-3.65400	13.35172	6.74967	3.41215
4	17.00	4.00	68.0000	16.0000	1.65280	6.84600	46.86772	11.31507	2.73175
5	8.50	1.33	11.3050	1.7689	-1.01720	-1.65400	2.73572	1.68245	1.03470
6	8.80	1.00	8.8000	1.0000	-1.34720	-1.35400	1.83332	1.82411	1.81495
7	22.00	6.17	135.7400	38.0689	3.82280	11.84600	140.32772	45.28489	14.61380
8	13.00	3.42	44.4600	11.6964	1.07280	2.84600	8.09972	3.05319	1.15090
9	12.50	3.67	45.8750	13.4689	1.32280	2.34600	5.50372	3.10329	1.74980
10	15.50	5.42	84.0100	29.3764	3.07280	5.34600	28.57972	16.42719	9.44210
11	9.50	1.17	11.1150	1.3689	-1.17720	-0.65400	0.42772	0.76989	1.38580
12	15.50	4.42	68.5100	19.5364	2.07280	5.34600	28.57972	11.08119	4.29650
13	9.50	1.17	11.1150	1.3689	-1.17720	-0.65400	0.42772	0.76989	1.38580
14	14.50	2.75	39.8750	7.5625	0.40280	4.34600	18.88772	1.75057	0.16225
15	19.00	6.25	118.7500	39.0625	3.90280	8.84600	78.25172	34.52417	15.23185
16	9.00	1.50	13.5000	2.2500	-0.84720	-1.15400	1.33172	0.97767	0.71775
17	14.00	4.25	59.5000	18.0625	1.90280	3.84600	14.79172	7.31817	3.62065
18	10.50	2.00	21.0000	4.0000	-0.34720	0.34600	0.11972	-0.12013	0.12055
19	6.00	0.42	2.5200	0.1764	-1.92720	-4.15400	17.25572	8.00559	3.71410
20	15.00	5.58	83.7000	31.1364	3.23280	4.84600	23.48372	15.66615	10.45100
21	13.00	3.42	44.4600	11.6964	1.07280	2.84600	8.09972	3.05319	1.15090
22	21.00	6.17	129.5700	38.0689	3.82280	10.84600	117.63572	41.46209	14.61380
23	12.00	3.00	36.0000	9.0000	0.65280	1.84600	3.40772	1.20507	0.42615
24	17.50	5.25	91.8750	27.5625	2.90280	7.34600	53.96372	21.32397	8.42625
25	5.50	0.33	1.8150	0.1089	-2.01720	-4.65400	21.65972	9.38805	4.06910
26	5.30	0.33	1.7490	0.1089	-2.01720	-4.85400	23.56132	9.79149	4.06910
27	6.50	0.75	4.8750	0.5625	-1.59720	-3.65400	13.35172	5.83617	2.55105
28	13.50	3.83	51.7050	14.6689	1.48280	3.34600	11.19572	4.96145	2.19870
29	4.50	0.25	1.1250	0.0625	-2.09720	-5.65400	31.96772	11.85757	4.39825
30	15.50	4.75	73.6250	22.5625	2.40280	5.34600	28.57972	12.84537	5.77345
31	16.50	4.67	77.0550	21.8089	2.32280	6.34600	40.27172	14.74049	5.39540
32	11.00	1.75	19.2500	3.0625	-0.59720	0.84600	0.71572	-0.50523	0.35665
33	17.50	5.25	91.8750	27.5625	2.90280	7.34600	53.96372	21.32397	8.42625
34	14.55	4.83	70.2765	23.3289	2.48280	4.39600	19.32482	10.91439	6.16430
35	10.00	2.00	20.0000	4.0000	-0.34720	-0.15400	0.02372	0.05347	0.12055
36	4.00	0.17	0.6800	0.0289	-2.17720	-6.15400	37.87172	13.39849	4.74020
37	3.50	0.08	0.2800	0.0064	-2.26720	-6.65400	44.27572	15.08595	5.14020
38	8.00	1.00	8.0000	1.0000	-1.34720	-2.15400	4.63972	2.90187	1.81495
39	8.00	1.33	10.6400	1.7689	-1.01720	-2.15400	4.63972	2.19105	1.03470
40	14.00	3.75	52.5000	14.0625	1.40280	3.84600	14.79172	5.39517	1.96785
41	1.75	0.17	0.2975	0.0289	-2.17720	-8.40400	70.62722	18.29719	4.74020
42	3.20	0.08	0.2560	0.0064	-2.26720	-6.95400	48.35812	15.76611	5.14020
43	5.55	0.33	1.8315	0.1089	-2.01720	-4.60400	21.19682	9.28719	4.06910

المشاهدة	y_i	x_i	$x_i y_i$	x_i^2	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(y_i-\bar{y})(x_i-\bar{x})$	$(x_i-\bar{x})^2$
44	2.75	0.08	0.2200	0.0064	-2.26720	-7.40400	54.81922	16.78635	5.14020
45	1.35	0.01	0.0135	0.0001	-2.33720	-8.80400	77.51042	20.57671	5.46250
46	5.50	0.58	3.1900	0.3364	-1.76720	-4.65400	21.65972	8.22455	3.12300
47	4.50	0.08	0.3600	0.0064	-2.26720	-5.65400	31.96772	12.81875	5.14020
48	3.25	0.02	0.0650	0.0004	-2.32720	-6.90400	47.66522	16.06699	5.41586
49	3.30	0.00	0.0000	0.0000	-2.34720	-6.85400	46.97732	16.08771	5.50935
50	1.40	0.08	0.1120	0.0064	-2.26720	-8.75400	76.63252	19.84707	5.14020
المجموع	507.70	117.36	1739.226	491.6858	0.0000	0.0000	1498.1992	547.5526	216.2184

٢-٧ خصائص مقدرات المربعات الصغرى: نظرية جاوس-ماركوف:

تنص نظرية جاوس-ماركوف على التالي:

"باستيفاء اشتراطات نموذج الانحدار، تتميز مقدرات المربعات الصغرى الاعتيادية بأنها أفضل مقدرات خطية غير متحيزة ((Best Linear Unbiased Estimators (BLUE))". وطبقًا لهذه النظرية تتصف مقدرات المربعات الصغرى بالخصائص التالي:

٢-٧-١ الخطية (Linearity): أي أن مقدرات المربعات الصغرى دوال خطية للمشاهدات الفعلية للمتغير التابع (Y_i).

البرهان:

معامل الانحدار $\hat{\beta}_1$:

يمكن كتابة المعادلة (2.20) كما يلي:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{S_{xx}}$$

ويمكن إعادة كتابة المعادلة أعلاه كما يلي:

$$\hat{\beta}_1 = \sum_{i=1}^{n} k_i y_i \qquad (2\text{-}25)$$

حيث إن $k_i = \frac{(x_i-\bar{x})}{S_{xx}}$

ويلاحظ من المعادلة (2.25) أن المعامل $\hat{\beta}_1$ دالة خطية للمتغير Y_i، أي أن $\hat{\beta}_1$ مجموع مرجح لقيم المتغير التابع حيث تعتبر k_i أوزانًا مرجحة، أي بطريقة أخرى فإن:

$$\hat{\beta}_1 = k_1Y_1 + k_2Y_2 + \dots + k_nY_n$$

ومن خصائص الأوزان (k_i) التي يمكن الاستفادة منها في البراهين اللاحقة نذكر ما يلي:

$$\sum_{i=1}^{n} k_i = 0 \ , \ \sum_{i=1}^{n} k_i^2 = \frac{1}{S_{xx}} \ , \ \sum_{i=1}^{n} k_i\left(x_i - \bar{x}\right) = \sum_{i=1}^{n} k_i x_i = 1$$

المعامل الثابت $\hat{\beta}_0$:

وباتباع نفس الطريقة التي اتُّبِعت في إثبات خطية المعامل $\hat{\beta}_1$ للمتغير التابع يمكننا إثبات خطية المعامل $\hat{\beta}_0$. وباستخدام المعادلة (2.25) يمكن كتابة المعادلة (2.19) كما يلي:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \tfrac{1}{n}\sum_{i=1}^{n} y_i - (\sum_{i=1}^{n} k_i y_i)\bar{x} = \sum_{i=1}^{n}(\frac{1}{n} - \bar{x}k_i)y_i \qquad (2\text{-}26)$$

وبذلك يكون $\hat{\beta}_0$ دالة خطية لقيم المتغير التابع Y_i لأن القوس في المعادلة (2.26) يحتوي على ثوابت هي ($k_i, \bar{x}, n$).

٢-٧-٢ عدم التحيز (Unbiasedness):

يسمى المقدر مقدراً غير متحيزٍ إذا كانت القيمة المتوسطة للمقدر مساوية للقيمة الحقيقية للمعلمة. وفيما يلي إثبات عدم تحيز المقدرين $\hat{\beta}_0$ و $\hat{\beta}_1$

عدم تحيز المقدر $\hat{\beta}_1$:

باستخدام المعادلة (2.4) يمكن كتابة المعادلة (2.25) كما يلي:

$$\hat{\beta}_1 = \sum_{i=1}^{n} k_i y_i = \sum_{i=1}^{n} k_i(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 \sum_{i=1}^{n} k_i + \beta_1 \sum_{i=1}^{n} k_i x_i + \sum_{i=1}^{n} k_i \varepsilon_i$$

وباستخدام خصائص الأوزان k_i نجد أن هذه المعادلة تأخذ الصيغة التالية:

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} k_i \varepsilon_i \qquad (2\text{-}27)$$

وبأخذ التوقع لكل من طرفي المعادلة نجد أن:

$$E(\hat{\beta}_1) = E(\beta_1 + \sum_{i=1}^{n} k_i \varepsilon_i) = E(\beta_1) + \sum_{i=1}^{n} k_i E(\varepsilon_i) = \beta_1$$

وذلك لأن توقع الثابت هو الثابت نفسه، وتوقع الثابت في متغير هو الثابت ضرب توقع المتغير، ونتيجة للفرض ($E(\varepsilon_i) = 0$) نحصل على النتيجة أعلاه. ومن ثم نستنتج أن $\hat{\beta}_1$ مقدر غير متحيز للمعلمة β_1.

عدم تحيز المقدر $\hat{\beta}_0$:

باستخدام المعادلة (2.4) يمكن كتابة المعادلة (2.26) كما يلي:

$$\hat{\beta}_0=\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}k_i\right)y_i=\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}k_i\right)(\beta_0+\beta_1x_i+\varepsilon_i)$$
$$\hat{\beta}_0=\sum_{i=1}^{n}\left(\frac{\beta_0}{n}-\beta_0\bar{x}k_i+\frac{\beta_1x_i}{n}-\beta_1\bar{x}k_ix_i+\frac{\varepsilon_i}{n}-\bar{x}k_i\varepsilon_i\right)$$
$$\hat{\beta}_0=\beta_0-\bar{x}\sum_{i=1}^{n}k_i\varepsilon_i$$

وبأخذ التوقع لكل من طرفي المعادلة نجد أن:

$$E(\hat{\beta}_0)=E(\beta_0-\bar{x}\sum_{i=1}^{n}k_i\varepsilon_i)=E(\beta_0)-\bar{x}\sum_{i=1}^{n}k_iE(\varepsilon_i)=\beta_0$$

وبذلك يعتبر المقدر $\hat{\beta}_0$ غير متحيز للمعلمة الحقيقية β_0. والخلاصة أن مقدرات المربعات الصغرى غير متحيزة.

٢-٧-٣ خاصية أقل تباين يمكن الحصول عليه (Minimum variance) أو خاصية الكفاءة (Efficiency):

إذا كان لدينا أكثر من مقدر غير متحيز لمعلمة ما وكان تباين أحدهما أقل من تباينات المقدرات الأخرى، يعتبر المقدر ذو التباين الأقل أكفأ من بقية المقدرات. وتنص نظرية جاوس-ماركوف على أن مقدرات المربعات الصغرى كفؤة. وقبل أن نبرهن هذه الخاصية لا بد من معرفة صيغ تباين هذه المقدرات.

تباين المقدر $\hat{\beta}_1$:

ومن تعريف التباين فان تباين المقدر $\hat{\beta}_1$ يأخذ الصيغة التالية:

$$var(\hat{\beta}_1)=E\left[\hat{\beta}_1-E(\hat{\beta}_1)\right]^2$$

وبما أن $E(\hat{\beta}_1)=\beta_1$ فإن:

$$var(\hat{\beta}_1)=E\left[\hat{\beta}_1-\beta_1\right]^2$$

وباستخدام المعادلة (2.27) يمكن كتابة معادلة تباين المقدر $\hat{\beta}_1$ كما يلي:

$$\begin{aligned}var(\hat{\beta}_1)&=E(\sum_{i=1}^{n}k_i\varepsilon_i)^2\\&=E(k_1^2\varepsilon_1^2+k_2^2\varepsilon_2^2+...+k_n^2\varepsilon_n^2+k_1k_2\varepsilon_1\varepsilon_2+...+k_{n-1}k_n\varepsilon_{n-1}\varepsilon_n)\\&=k_1^2\sigma^2+k_2^2\sigma^2+...+k_n^2\sigma^2+0+0+...+0\\&=\sigma^2\sum_{i=1}^{n}k_i^2=\frac{\sigma^2}{S_{xx}}\end{aligned}\qquad(2.28)$$

حيث تم استخدام اشتراط استقلال حدود الخطأ $\left\{E\left(\varepsilon_i\varepsilon_j\right)=0\ ,\ i\neq j\right\}$ وثبات تباين حـد الخطأ $E(\varepsilon_i^2)=\sigma^2$ في البرهان. ويمكن كتابة الانحراف المعياري للمقدر $\hat{\beta}_1$ كما يلي:

$$\text{std}(\hat{\beta}_1)=\frac{\sigma}{\sqrt{S_{xx}}} \qquad (2\text{-}29)$$

تباين المقدر $\hat{\beta}_0$:

ومن تعريف التباين فإن تباين المقدر $\hat{\beta}_0$ يأخذ الصيغة التالية:

$$\text{var}\left(\hat{\beta}_0\right) = E\left[\hat{\beta}_0-E\left(\hat{\beta}_0\right)\right]^2$$

وبما أن $E(\hat{\beta}_0)=\beta_0$ فإن:

$$\text{var}\left(\hat{\beta}_0\right) = E\left[\hat{\beta}_0-\beta_0\right]^2$$

وباستخدام المعادلة (2.26) يمكن كتابة تباين $\hat{\beta}_0$ كما يلي:

$$\begin{aligned}\text{var}(\hat{\beta}_0)&=E\left[\sum_{i=1}^{n}(\tfrac{1}{n}-\bar{x}k_i)\varepsilon_i\right]^2=\sum_{i=1}^{n}\left(\tfrac{1}{n}-\bar{x}k_i\right)^2E\left(\varepsilon_i\right)^2\\&=\sigma^2\sum_{i=1}^{n}\left(\tfrac{1}{n}-\bar{x}k_i\right)^2=\sigma^2\left(\tfrac{S_{xx}+n\bar{x}^2}{nS_{xx}}\right)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{nS_{xx}}\end{aligned} \qquad (2\text{-}30)$$

أما الانحراف المعياري للمقدر $\hat{\beta}_0$ فهو:

$$\text{std}(\hat{\beta}_0)=\sqrt{\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{nS_{xx}}} \qquad (2\text{-}31)$$

التغاير بين المعاملين $\hat{\beta}_0$ و $\hat{\beta}_1$:

يعرف التغاير بين $\hat{\beta}_0$ و $\hat{\beta}_1$ كما يلي:

$$\text{cov}\left(\hat{\beta}_0,\hat{\beta}_1\right) = E\left[\left(\hat{\beta}_0-E\left(\hat{\beta}_0\right)\right)\left(\hat{\beta}_1-E\left(\hat{\beta}_1\right)\right)\right]$$

أي

$$\text{cov}\left(\hat{\beta}_0,\hat{\beta}_1\right) = E\left[\left(\hat{\beta}_0-\beta_0\right)\left(\hat{\beta}_1-\beta_1\right)\right]$$

بما أن:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

وبأخذ التوقع لطرفي هذه المعادلة نحصل على:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

وبطرح هذه المعادلة من سابقتها نحصل على:

$$\hat{\beta}_0 - \beta_0 = -\bar{x}(\hat{\beta}_1 - \beta_1)$$

والآن بالتعويض في معادلة التغاير نحصل على:

$$\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = E\left(-\bar{x}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1)\right) = -\bar{x}E(\hat{\beta}_1 - \beta_1)^2 = -\bar{x}\,\text{var}(\hat{\beta}_1) = -\bar{x}\,\frac{\sigma^2}{S_{xx}}$$

(2-32)

والآن بعد تحديد صيغ تباين معاملي الانحدار يمكننا إثبات خاصية أقل تباين باستخدام معادلة شبيهة بالمعادلة (2.25) بتعريف مقدر خطي آخر كما يلي:

$$D = \sum_{i=1}^{n} h_i y_i \qquad (2\text{-}33)$$

حيث إن D المقدر الجديد للمعلمة $\hat{\beta}_1$ وh_i أوزان مرجحة جديدة.

ويمكن إعادة كتابة المعادلة (2.33) كما يلي:

$$D = \sum_{i=1}^{n} h_i(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 \sum_{i=1}^{n} h_i + \beta_1 \sum_{i=1}^{n} h_i x_i + \sum_{i=1}^{n} h_i \varepsilon_i$$

ويعتبر D مقدر غير متحيز للمعلمة β_1 إذا تحقق الشرطان التاليان:

$$\sum_{i=1}^{n} h_i x_i = 1 \quad \text{و} \quad \sum_{i=1}^{n} h_i = 0$$

والآن يمكن كتابة تباين المقدر D كما يلي:

$$\text{var}(D) = \text{var}\left(\sum_{i=1}^{n} h_i Y_i\right) = \sum_{i=1}^{n} h_i^2 \,\text{var}(Y_i)$$

وبما أن

$$\text{var}(\varepsilon_i) = \text{var}(Y) = \sigma^2$$

فإن:

$$\operatorname{var}(D)=\sigma^2\sum_{i=1}^{n}h_i^2$$

وبإضافة وطرح القيمة $\frac{(x_i-\bar{x})}{S_{xx}}$ للحد h_i في المعادلة أعلاه، نحصل على

$$\operatorname{var}(D)=\sigma^2\left(\sum_{i=1}^{n}\left(h_i-\frac{(x_i-\bar{x})}{S_{xx}}+\frac{(x_i-\bar{x})}{S_{xx}}\right)^2\right)$$

$$\operatorname{var}(D)=\sigma^2\left(\sum_{i=1}^{n}\left(h_i-\frac{(x_i-\bar{x})}{S_{xx}}\right)^2+\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(S_{xx})^2}+2\sum_{i=1}^{n}\left(h_i-\frac{(x_i-\bar{x})}{S_{xx}}\right)\left(\frac{(x_i-\bar{x})}{S_{xx}}\right)\right)$$

$$\operatorname{var}(D)=\sigma^2\sum_{i=1}^{n}\left(h_i-\frac{(x_i-\bar{x})}{S_{xx}}\right)^2+\frac{\sigma^2}{S_{xx}}$$

(2.34)

ويلاحظ من المعادلة (2.34) أن تباين المقدر D يساوي تباين المقدر $\hat{\beta}_1$ إذا تساوت الأوزان k_i و h_i $(h_i=k_i)$، أما إذا اختلفت الأوزان سيبقى تباين المقدر $\hat{\beta}_1$ أقل من تباين المقدر D وبالتالي تعتبر مقدرات المربعات الصغرى مقدرات ذات كفاءة.

٢-٨ تقدير التباين (σ^2):

يتم قياس الانحرافات أو الفروق بين قيم المتغير المُشاهدة وقيمها المقدرة بحساب تباين البواقي (S^2) كمقدر لتباين حد الخطأ العشوائي σ^2. ويقيس التباين مدى تشتت قيم المتغير التابع الفعلية حول خط الانحدار المقدر، فكلما زادت قيمة التباين دلّ ذلك على زيادة درجة تشتت القيم المشاهدة حول خط الانحدار والعكس صحيح. ورياضياً يعرف التباين كما يلي:

$$S^2=\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{n-2}=\frac{\sum_{i=1}^{n}e_i^2}{n-2} \qquad (2\text{-}35)$$

كما أن تباين البواقي مقدر غير متحيز لتباين حد الخطأ العشوائي، أي أن:

$$E(S^2)=\sigma^2$$

إثبات عدم تحيز مقدر التباين:

بما أن:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

وبجمع وقسمة طرفي المعادلة على n نحصل على:

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}$$

وبطرح هذه المعادلة عن سابقتها نحصل على:

$$y_i - \bar{y} = \beta_1 (x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})$$

وبما أن

$$e_i = \left(y_i - \bar{y}\right) - \hat{\beta}_1 \left(x_i - \bar{x}\right)$$

فإنه بالتعويض عن قيمة $\left(y_i - \bar{y}\right)$ نحصل على:

$$e_i = \beta_1 \left(x_i - \bar{x}\right) + (\varepsilon_i - \bar{\varepsilon}) - \hat{\beta}_1 \left(x_i - \bar{x}\right)$$
$$e_i = -(\hat{\beta}_1 - \beta)\left(x_i - \bar{x}\right)^2 + (\varepsilon_i - \bar{\varepsilon})$$

وبتربيع طرفي المعادلة ومن ثم جمعهما نحصل على:

$$\sum_{i=1}^{n} e_i = (\hat{\beta}_1 - \beta_1)^2 S_{xx} + \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2 - 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})\left(x_i - \bar{x}\right)$$
$$= (\hat{\beta}_1 - \beta_1)^2 S_{xx} + \sum_{i=1}^{n} \varepsilon_i^2 - n\bar{\varepsilon}^2 - 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{n} \varepsilon_i x_i$$

وبأخذ التوقع لطرفي المعادلة نحصل على:

$$E\left(\sum_{i=1}^{n} e_i\right) = E(\hat{\beta}_1 - \beta_1)^2 S_{xx} + E\left(\sum_{i=1}^{n} \varepsilon_i^2\right) - E\left(n\bar{\varepsilon}^2\right) - 2E(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{n} \varepsilon_i x_i$$
$$= \sigma^2 \qquad + n\sigma^2 \qquad - \sigma^2 \qquad - 2E\left(\sum_{i=1}^{n} k_i \varepsilon_i \sum_{i=1}^{n} \varepsilon_i x_i\right)$$
$$= \sigma^2 \qquad + n\sigma^2 \qquad - \sigma^2 \qquad - 2\sigma^2$$

وبقسمة طرفي المعادلة على (n-2) نحصل على القيمة المقدرة للتباين (S^2) كما يلي:

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$$

حيث تمت الاستفادة من النتائج التالية:

$$E(\hat{\beta}_1 - \beta_1)^2 = \frac{\sigma^2}{S_{xx}} \Rightarrow \sigma^2 = E(\hat{\beta}_1 - \beta)^2 S_{xx}$$

$$E(\hat{\beta}_1 - \beta_1) = \sum_{i=1}^{n} k_i \varepsilon_i$$

$$E\left(\sum_{i=1}^{n} k_i \varepsilon_i\right)^2 = E\left(\frac{\left(\sum_{i=1}^{n} x_i \varepsilon_i\right)^2}{(s_{xx})^2}\right) = \frac{\sigma^2}{s_{xx}} \Rightarrow \sigma^2 = \frac{\left(\sum_{i=1}^{n} x_i \varepsilon_i\right)^2}{s_{xx}}$$

ويمكن إيجاد القيمة المتوقعة لـ S^2 كما يلي:

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n} e_i^2}{n-2}\right) = \frac{(n-2)\sigma^2}{n-2} = \sigma^2$$

وبذلك يكون S^2 مقدراً غير متحيزاً للمعلمة σ^2. وبما أن قيمة التباين مجهولة، ستتبدل σ^2 الواردة في معادلات التباين والأخطاء المعيارية لمقدرات المربعات الصغرى بـ S^2 وعلى ذلك تأخذ تقديرات الأخطاء المعيارية لهذه المقدرات الصيغ التالية:

الخطأ المعياري لـ $\hat{\beta}_1$:

$$s.e(\hat{\beta}_1) = \frac{S}{\sqrt{S_{xx}}} \qquad (2\text{-}36)$$

الخطأ المعياري لـ $\hat{\beta}_0$:

$$s.e(\hat{\beta}_0) = \sqrt{\frac{S^2 \sum_{i=1}^{n} x_i^2}{n \times S_{xx}}} \qquad (2\text{-}37)$$

تغاير المقدرين $\hat{\beta}_1$ و $\hat{\beta}_0$:

$$\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x} \times S^2}{S_{xx}} \qquad (2\text{-}38)$$

مثال:

من بيانات نموذج انحدار وزن الطفل احسب الخطأ المعياري (S) والخطأ المعياري لمعاملي الانحدار $\hat{\beta}_0$ و $\hat{\beta}_1$ وتغاير $\hat{\beta}_0$ و $\hat{\beta}_1$.

الحل:

- الخطأ المعياري (S):

يوضح الجدول رقم (٢-٤) الحسابات المطلوبة لإيجاد الخطأ المعياري. وبأخذ الجذر التربيعي للتباين (المعادلة (2.35)) نحصل على قيمة الخطأ المعياري كما يلي:

$$S=\sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}=\sqrt{\frac{111.5744}{(50-2)}}=1.525$$

- الخطأ المعياري لـ $\hat{\beta}_1$:

باستخدام المعادلة (2.36) يتم حساب الخطأ المعياري $\hat{\beta}_1$ كما يلي:

$$s.e(\beta_1)=\sqrt{\frac{S^2}{S_{xx}}}=\sqrt{\frac{2.325}{216.218}}=0.104$$

- الخطأ المعياري لـ $\hat{\beta}_0$:

باستخدام المعادلة (2.37) يتم حساب الخطأ المعياري $\hat{\beta}_0$ كما يلي:

$$s.e(\beta_0)=\sqrt{\frac{S^2\sum_{i=1}^{n} x_i^2}{n\times S_{xx}}}=\sqrt{\frac{2.325\times 491.6858}{50\times 216.218408}}=0.325$$

- تغاير $\hat{\beta}_0$ و $\hat{\beta}_1$:

باستخدام المعادلة (2.38) يتم حساب تغاير $\hat{\beta}_0$ و $\hat{\beta}_1$ كما يلي:

$$\text{cov}(\beta_0,\beta_1)=\frac{-\bar{x}S^2}{S_{xx}}=\frac{-2.3472\times 2.325}{216.218408}=-0.0252$$

جدول رقم (٢-٤): الحسابات اللازمة لحساب الأخطاء المعيارية لمقدرات المربعات الصغرى

المشاهدة	$\hat{y}_i = \beta_0 + \beta_1 x_i$	$e_i = y_i - \hat{y}_i$	e_i^2	x_i^2
1	11.80715	-0.30715	0.09434	9.00000
2	16.87196	-0.87196	0.76032	25.00000
3	5.47614	1.02386	1.04829	0.25000
4	14.33956	2.66044	7.07795	16.00000
5	7.57804	0.92196	0.85001	1.76890
6	6.74234	2.05766	4.23395	1.00000
7	19.83488	2.16512	4.68776	38.06890
8	12.87076	0.12924	0.01670	11.69640
9	13.50387	-1.00387	1.00775	13.46890
10	17.93557	-2.43557	5.93202	29.37640
11	7.17285	2.32715	5.41561	1.36890
12	15.40317	0.09683	0.00938	19.53640
13	7.17285	2.32715	5.41561	1.36890
14	11.17405	3.32595	11.06193	7.56250
15	20.03747	-1.03747	1.07634	39.06250
16	8.00855	0.99145	0.98298	2.25000
17	14.97266	-0.97266	0.94607	18.06250
18	9.27475	1.22525	1.50124	4.00000
19	5.27355	0.72645	0.52773	0.17640
20	18.34076	-3.34076	11.16067	31.13640
21	12.87076	0.12924	0.01670	11.69640
22	19.83488	1.16512	1.35751	38.06890
23	11.80715	0.19285	0.03719	9.00000
24	17.50506	-0.00506	0.00003	27.56250
25	5.04563	0.45437	0.20645	0.10890
26	5.04563	0.25437	0.06470	0.10890
27	6.10924	0.39076	0.15269	0.56250
28	13.90905	-0.40905	0.16732	14.66890
29	4.84304	-0.34304	0.11768	0.06250
30	16.23886	-0.73886	0.54592	22.56250
31	16.03627	0.46373	0.21505	21.80890
32	8.64165	2.35835	5.56183	3.06250
33	17.50506	-0.00506	0.00003	27.56250
34	16.44145	-1.89145	3.57760	23.32890
35	9.27475	0.72525	0.52599	4.00000
36	4.64045	-0.64045	0.41017	0.02890
37	4.41253	-0.91253	0.83271	0.00640
38	6.74234	1.25766	1.58170	1.00000
39	7.57804	0.42196	0.17805	1.76890
40	13.70646	0.29354	0.08617	14.06250
41	4.64045	-2.89045	8.35469	0.02890
42	4.41253	-1.21253	1.47023	0.00640
43	5.04563	0.50437	0.25439	0.10890
44	4.41253	-1.66253	2.76401	0.00640
45	4.23526	-2.88526	8.32474	0.00010
46	5.67873	-0.17873	0.03195	0.33640
47	4.41253	0.08747	0.00765	0.00640
48	4.26059	-1.01059	1.02129	0.00040
49	4.20994	-0.90994	0.82799	0.00000
50	4.41253	-3.01253	9.07535	0.00640
المجموع	507.7000	0.0000	111.5744	491.6858

٢-٩ الاستدلال الإحصائي (Statistical Inference):

٢-٩-١ اختبار الفروض (Hypothesis Testing):

ذكرنا في الفصل الأول أن من أهم مراحل بناء النموذج وضع أو صياغة فروض البحث. والفرض (Hypothesis) في تحليل الانحدار هو قيم مبدئية يضعها الباحث لمعالم نموذج الانحدار، حيث يُجرى التحديد المبدئي لقيم وإشارات معاملات نموذج الانحدار. وتصاغ الفروض عادة بعد مراجعة أدبيات المشكلة التي تتضمن النظريات، والدراسات والبحوث السابقة والملاحظة. وفي الإحصاء يطلق على الفرض الذي يتم وضعه **بفرض العدم** (Null Hypothesis) ويرمز له بـ 'H_0' حيث يفترض الباحث عادة عدم وجود فرق بين قيمة المعلمة الحقيقية وقيمة محددة؛ أو عدم وجود اختلاف بين قيمتي معلمتين، ولذلك جاءت التسمية بفرض العدم أو الفرض الصفري. ويصاحب الفرض العدم عادة **فرض بديل** (Alternative Hypothesis) يرمز له بـ H_1، وهو الفرض الذي يقبل إذا رفض فرض العدم. وينقسم الفرض البديل إلى **فرض بسيط** (Simple hypothesis) و**مركب** (Composite Hypothesis). مثلًا عندما يساوي الفرض البديل قيمة محددة يسمى بالفرض البسيط مثل $H_1{:}\beta_1{=}0.7$، وعند وضع الفرض البديل في صيغة "إن قيمة المعلمة لا تختلف عن الصفر أو لأي قيمة أخرى" بغض النظر عن كون الاختلاف موجبًا أو سالبًا مثل $H_1{:}\beta_1 \neq 0.7$ يعرف الفرض بالفرض المركب. ومن الممكن صياغة فرض العدم في اتجاه واحد (One-Tailed) إذا توافرت لدى الباحث معلومات مسبقة مبنية على عمل تجريبي وبالتالي يكون الفرض البديل أيضًا ذا طرف واحد. المثال التالي يوضح فرض العدم ذا الطرف الواحد:

$$H_0{:}\beta_1 \leq \beta_1^* \quad \text{مقابل} \quad H_1{:}\ \beta_1 > \beta_1^*$$

أو

$$H_0:\beta_1 \geq \beta_1^* \quad \text{مقابل} \quad H_1{:}\ \beta_1 < \beta_1^*$$

حيث إن β_1^* القيمة الفّرضية للمعلمة β_1.

اختبار المعنوية:

اختبار المعنوية هو إجراء يستخدم نتائج العينة للتأكد من صحة أو خطأ فرض العدم. وإن قرار رفض أو قبول فرض العدم يعتمد على إحصائية الاختبار (Test Statistic) التي يتم الحصول على قيمتها من بيانات العينة. وإحصائية الاختبار هي متغير عشوائي له توزيع احتمالي يصف العلاقة بين القيمة النظرية والقيم المشاهدة من العينة. ويتم عادة مقارنة قيمة إحصائية الاختبار المحسوبة من بيانات العينة مع قيمته النظرية، وعليه يتم اتخاذ قرار رفض أو عدم رفض فرض العدم. وعند اتخاذ القرار حول فرض العدم يوجد نوعان من الأخطاء يمكن الوقوع فيهما؛ إذا رُفض فرض العدم الصحيح يطلق عليه **خطأ من النوع الأول** (Type I error) ويرمز لاحتمال هذا الخطأ بـ α ويسمى بمستوى المعنوية أو مستوى الدلالة، كما أن عدم رفض فرض العدم غير الصحيح يطلق عليه **خطأ من النوع الثاني** (Type II error)، ويرمز لاحتمال هذا الخطأ بـ β. وعمومًا توجد أربعة قرارات ممكنة عند اتخاذ القرار حول فرض العدم هي:

١- رفض فرض العدم وهو صحيح ويطلق عليه خطأ من **النوع الأول** وتمثل α احتمال وقوع هذا الخطأ.

٢- عدم رفض فرض العدم وهو صحيح. وهذا قرار صحيح واحتمال اتخاذ هذا القرار هو $1-\alpha$.

٣- عدم رفض فرض العدم وهو خاطئ ويسمى هذا الخطأ **النوع الثاني** واحتمال الوقوع في هذا الخطأ هو β.

٤- رفض فرض العدم وهو خاطئ وهذا قرار سليم واحتمال اتخاذه هو $(1-\beta)$.

ويمكن التقليل من احتمالية الوقوع في الخطأ من النوع الأول بخفض مستوى المعنوية كاستخدام مستوى المعنوية ١% بدلًا من ٥% مثلًا ولكن يجب ملاحظة أن تقليل احتمال الوقوع في الخطأ من النوع الأول يزيد من احتمال الوقوع في الخطأ من النوع الثاني.

مستوى الدلالة (Level of Significance):

يعرف مستوى المعنوية أو مستوى الدلالة (α) على أنه الحد الأقصى لاحتمال وقوعنا في أخطاء من **النوع الأول** عند اختبارنا لفرض العدم. ويحدد عادة مستوى المعنوية قبل إجراء الاختبار وتأخذ القيمة ٥% أو ٠,٠٥ وأحيانًا بـ ١% و١٠%. (Osborn, 2006) وبالطبع يمكن أخذ أي قيمة أخرى غير المتعارف عليها كـ ٣%، ٤%. وتعني كلمة المعنوية "أن الفرق بين القيمة النظرية للمعلمة في المجتمع والقيمة الناتجة من العينة فرق حقيقي ولا يعزى إلى الصدفة". ومستوى المعنوية ٥% مثلًا يعني أنه إذا تم سحب (١٠٠) عينة من نفس المجتمع فمن المحتمل أن نرفض فرض العدم وهو في الواقع صحيح خمس مرات. أو بمعنى آخر نكون على درجة ثقة ٩٥% من أن النتائج التي نحصل عليها صحيحة. وكذلك يجري التفسير لمستويات المعنوية الأخرى بنفس الطريقة.

اختبار معنوية معاملات الانحدار (اختبار 't'):

إن اختبار المعنوية إجراء يستخدم نتائج العينة للتأكد من صحة أو خطأ الفرض الصفري. ويعتمد اختبار المعنوية على إحصائية الاختبار (Test Statistic) وتوزيع معاينتها.

وباستيفاء اشتراط التوزيع الطبيعي لحد الخطأ العشوائي، نجد أن المقدرين $\hat{\beta}_0$ و $\hat{\beta}_1$ يتبعان التوزيع الطبيعي أيضًا، أي أن:

$$\hat{\beta}_1 \sim N\left(\beta_1,\ \frac{\sigma^2}{S_{xx}}\right) \qquad (2\text{-}39)$$

و

$$\hat{\beta}_0 \sim N\left(\beta_0,\ \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \times S_{XX}}\right) \qquad (2\text{-}40)$$

ومن المعادلة (2.39) يمكن حساب القيمة المعيارية للمقدر $\hat{\beta}_1$ بالطريقة المعتادة كما يلي:

$$Z = \frac{\hat{\beta}_1 - \beta_1}{std(\hat{\beta}_1)} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\sqrt{S_{XX}}}{\sigma} \sim N(0, 1) \qquad (2\text{-}41)$$

حيث إن $std\left(\hat{\beta}_1\right)$ هو الانحراف المعياري للمقدر $\hat{\beta}_1$ حسب المعادلة (2.29). وبما أن القيمة Q تتبع توزيـع تربيـع كاي (χ^2) بدرجات حرية (n-2) حيث

$$Q = \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2} \qquad (2\text{-}42)$$

فإن القيمة T حيث

$$T = \frac{Z}{\sqrt{Q/(n-2)}} \sim t_{n-2} \qquad (2\text{-}43)$$

لها توزيع ‘t’ بدرجات حرية (n-2).* وباستخدام مقدر الانحراف المعياري (S) يمكن إعادة كتابة هذه المعادلـة كـما يلي:

$$T = \frac{\hat{\beta}_1 - \beta_1}{s.e(\hat{\beta}_1)} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\sqrt{S_{XX}}}{S} \sim t_{n\text{-}2} \qquad (2.44)$$

وإذا ما تم تحديد قيمة المعلمة الحقيقية β_1 عند فرض العدم بـ β_1^*، يمكن حساب قيمة T من العينـة وبالتـالي يـتم استخدامها كإحصائية للاختبار. وبما أن الإحصائية T لها توزيع ‘t’ يمكننا إيجاد فترة ثقة لـ $\hat{\beta}_1$ كما يلي:

$$Pr[-t_{n-2,\alpha/2} \le \frac{\hat{\beta}_1 - \beta_1^*}{s.e(\hat{\beta}_1)} \le t_{n-2,\alpha/2}] = 1-\alpha \qquad (2\text{-}45)$$

حيث إن β_1^* قيمة المعلمة الفرضية (H_0)، و $t_{n-2,\alpha/2}$ القيمة الحرجة التي يتم استخراجها من جدول توزيع ‘t’ عنـد مستوى معنوية معين ودرجات حرية (n-2) و $s.e\left(\hat{\beta}_1\right)$ هو الخطأ المعياري للمقدر $\hat{\beta}_1$.

وبإعادة تنظيم المعادلة (2.45) يمكن الحصول على فترة الثقة التالية:

$$Pr[\beta_1^* - t_{n-2,\alpha/2} s.e(\hat{\beta}_1) \le \hat{\beta}_1 \le \beta_1^* + t_{n-2,\alpha/2} s.e(\hat{\beta}_1)] = 1-\alpha \qquad (2.46)$$

وتحتوي هذه الفترة على المقدر $\hat{\beta}_1$ باحتمال $1-\alpha$. وبلغة اختبار الفروض فـإن فـترة الثقـة التـي تـم الوصـول إليهـا حسب المعادلة (2.46) تعرف بمنطقة القبول (قبول فرض العدم) (Region of Acceptance) والمنـاطق التـي تقـع خـارج فترة الثقة تعرف بمناطق الرفض أو المناطق الحرجة (Regions of Rejection or Critical Regions) (انظر الشكل رقم (٢-٧)). فإذا وقعت قيمة الإحصائية T حسب المعادلة (2.44) في منطقة الرفض رفضنا فرض العدم عنـد مسـتوى معنويـة

* انظر الملحق رقم (أ) حول التوزيع الطبيعي المعياري، توزيع ‘t’ وتوزيع مربع كاي والعلاقة بينهما

α لصالح الفرض البديل، وأما إذا وقعت القيمة المشاهدة لإحصائية الاختبار في منطقة القبول فإننا نقبل الفرض الصفري. ويمكن تلخيص اختبارات المعنوية لمعاملي نموذج الانحدار الخطي البسيط فيما يلي:

اختبار معنوية المعلمة β_1:

- إحصائية الاختبار:

$$T=\frac{\hat{\beta}_1-\beta_1^*}{s.e(\hat{\beta}_1)}=\frac{\left(\hat{\beta}_1-\beta_1^*\right)\sqrt{S_{XX}}}{S} \tag{2-47}$$

- **قاعدة القرار:**

نوع الاختبار:	**فرض العدم (H_0):**	**الفرض البديل (H_1):**	**القرار*: نرفض H_0 إذا كان**
ذو طرفين	$\beta_1=\beta_1^*$	$\beta_1\neq\beta_1^*$	$\lvert T\rvert>t_{n-2,\alpha/2}$
ذو طرف واحد (أيمن)	$\beta_1\leq\beta_1^*$	$\beta_1>\beta_1^*$	$T>t_{n-2,\alpha}$
ذو طرف واحد (أيسر)	$\beta_1\geq\beta_1^*$	$\beta_1<\beta_1^*$	$T<-t_{n-2,\alpha}$

* $\lvert T\rvert$ القيمة المطلقة لـ T، المتباينات "أقل من" أو "أكبر من" مستخدمة كما هي بالإنجليزية، حيث "<" يرمز لأصغر من و">" لأكبر من

اختبار معنوية المعلمة β_0:

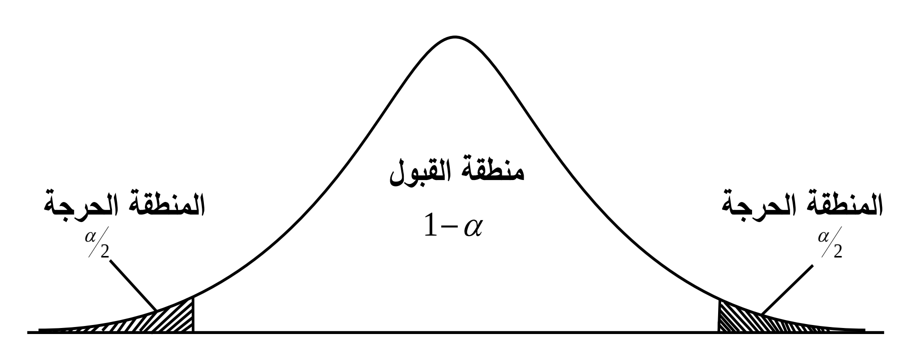
- إحصائية الاختبار:

$$T=\frac{\hat{\beta}_0-\beta_0^*}{s.e(\hat{\beta}_0)}=\frac{\left(\hat{\beta}_0-\beta_0^*\right)\sqrt{n\times S_{XX}}}{S\sqrt{\sum_{i=1}^{n}x_i^2}} \tag{2-48}$$

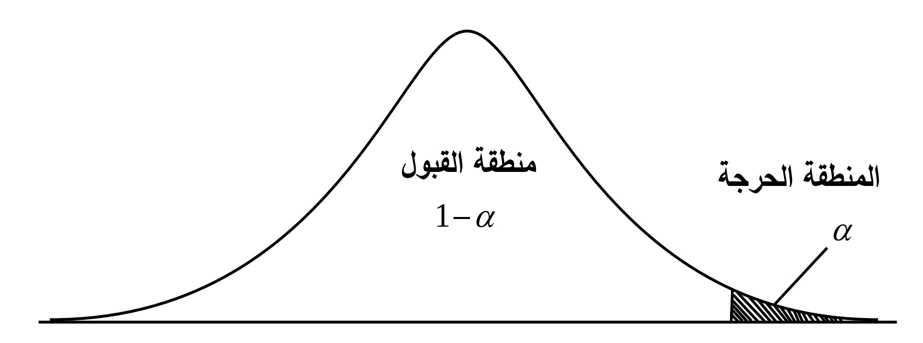
نوع الاختبار:	**فرض العدم (H_0):**	**الفرض البديل (H_1):**	**القرار*: نرفض H_0 إذا كان**
ذو طرفين	$\beta_0=\beta_0^*$	$\beta_0\neq\beta_0^*$	$\lvert T\rvert>t_{n-2,\alpha/2}$
ذو طرف واحد (أيمن)	$\beta_0\leq\beta_0^*$	$\beta_0>\beta_0^*$	$T>t_{n-2,\alpha}$
ذو طرف واحد (أيسر)	$\beta_0\geq\beta_0^*$	$\beta_0<\beta_0^*$	$T<-t_{n-2,\alpha}$

* $\lvert T\rvert$ القيمة المطلقة لـ T المتباينات "أقل من" أو "أكبر من" مستخدمة كما هي بالإنجليزية، حيث "<" يرمز لأصغر من و">" لأكبر من

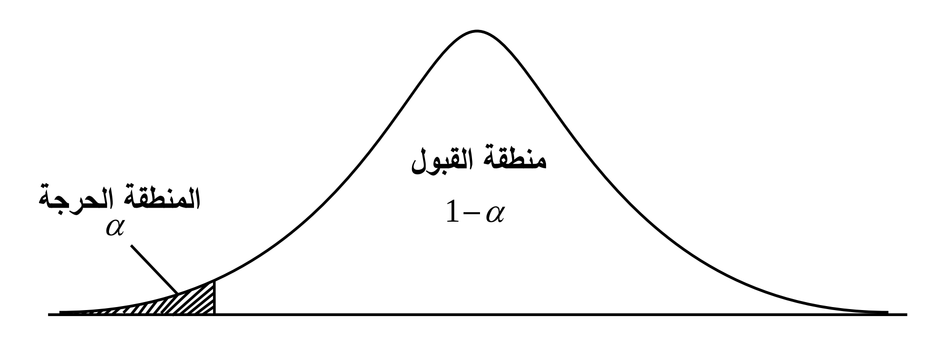
أ- اختبار ذو طرفين $H_1: \beta_1 \neq \beta_1^*$ مقابل $H_0: \beta_1 = \beta_1^*$



ب- اختبار ذو طرف واحد (أيمن) $H_1: \beta_1 > \beta_1^*$ مقابل $H_0: \beta_1 \leq \beta_1^*$



ج- اختبار ذو طرف واحد (أيسر) $H_1: \beta_1 < \beta_1^*$ مقابل $H_0: \beta_1 \geq \beta_1^*$



شكل (٢-٧): مناطق القبول والرفض (حالة المعلمة β_1)

٢-٩-٢ التقدير بفترة (Interval Estimation):

لقد ذُكر، فيما سبق، أن مقدرات المربعات الصغرى هي مقدرات نقطة (Point Estimators)، بمعنى أنها تعطي تقديراً واحداً لكل معلمة من معالم النموذج. وبالطبع يختلف تقدير المعلمة من عينة لأخرى. وفي الإحصاء يقاس مدى ثبات مقدرات النقطة بتباينها أو خطئها المعياري. ولذلك نحتاج إلى توفير معلومات حول مدى دقة التقديرات التي تؤخذ من العينات. فبدلًا من الاعتماد على قيمة واحدة نستطيع حساب احتمال لمدى معين يعرف **بفترة الثقة** يحتوي على قيمة المعلمة الحقيقية. ولحساب فترة الثقة يمكن إيجاد قيمتين موجبتين d و α بحيث يكون احتمال أن يحتوي المدى من $\left(\hat{\beta}_1 - d\right)$ إلى $\left(\hat{\beta}_1 + d\right)$ للمعلمة الحقيقية β_1 هو $(1-\alpha)$، أو ما يمكن التعبير عنه رياضياً كما يلي:

$$\Pr(\hat{\beta}_1 - d \le \beta_1 \le \hat{\beta}_1 + d) = 1 - \alpha \qquad (2\text{-}49)$$

حيث إن Pr يرمز للاحتمال و d قيمة ثابتة موجبة وتشير α إلى احتمال أن تقع المعلمة خارج حدود فترة الثقة (مستوى المعنوية)، وتُراوح قيمتها بين الصفر والواحد الصحيح. وهذا يعني أن هناك احتمال أن تقع المعلمة داخل حدود فترة الثقة $\left(\hat{\beta}_1 - d\right)$ إلى $\left(\hat{\beta}_1 + d\right)$. ويسمى المقدار $\left(\hat{\beta}_1 - d\right)$ بالحد الأدنى لفترة الثقة و $\left(\hat{\beta}_1 + d\right)$ بالحد الأعلى لها. وتنص المعادلة (2.49) على أنه في المعاينات المتكررة، إذا كان مستوى المعنوية يساوي ٥% مثلًا، فإن ٩٥ فترة من ١٠٠ فترة تحتوي على معلمة المجتمع المجهولة.

فترة الثقة لمعالم نموذج الانحدار البسيط:

من المعادلة (2.42) نعلم التوزيع الاحتمالي للمتغير T حيث

$$T = \frac{\hat{\beta}_1 - \beta_1}{s.e(\hat{\beta}_1)} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\sqrt{S_{XX}}}{S}$$

هو توزيع 't' بدرجات حرية (n-2) . وباستخدام توزيع t يتم إيجاد فترة الثقة للمعلمة β_1 كما يلي:

$$\Pr[-t_{n-2,\alpha/2} \le \frac{\hat{\beta}_1 - \beta_1}{s.e(\hat{\beta}_1)} \le t_{n-2,\alpha/2}] = 1 - \alpha$$

وبإعادة تنظيم هذه المعادلة نحصل على:

$$\Pr[\hat{\beta}_1 - t_{n-2,\alpha/2} s.e(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{n-2,\alpha/2} s.e(\hat{\beta}_1)] = 1 - \alpha \qquad (2\text{-}50)$$

حيث يشير (n-2) إلى درجات الحرية وn حجم العينة و2 عدد معالم نموذج الانحدار الخطي البسيط و $t_{n-2,\alpha/2}$ القيمة الحرجة التي يتم استخراجها من جدول توزيع 't' و $s.e\left(\hat{\beta}_1\right)$ هو الخطأ المعياري للمقدر $\hat{\beta}_1$. وعادة ما يكتفى بالصيغة التالية لفتر ثقة β_1:

$$\hat{\beta}_1 \pm t_{n-2,\alpha/2}\ s.e(\hat{\beta}_1) \qquad (2\text{-}51)$$

وباتباع نفس الطريقة يمكن إيجاد فترة الثقة للمعلمة β_0 كما يلي:

$$\hat{\beta}_0 \pm t_{n-2,\alpha/2}\ s.e(\hat{\beta}_0) \qquad (2\text{-}52)$$

ويلاحظ أن طول فترة الثقة لكل من المعلمتين تعتمد على التالي:

- تباين المقدر: أن طول فترة الثقة يزداد بزيادة تباين المقدر.
- عدد المشاهدات: كلما كان حجم العينة صغيراَ اتسعت فترة الثقة.
- مستوى المعنوية: كلما زاد مستوى المعنوية اتسعت فترة الثقة.

فترة الثقة لـ σ^2 :

كما سبق ذكره فإن للقيمة Q حسب المعادلة (2.42) توزيع مربع كاي بدرجات حرية $(n-2)$. وباستخدام توزيع مربع كاي يمكن إيجاد فترة ثقة لـ σ^2 كما يلي:

$$\Pr(\chi^2_{1-\alpha/2} \le Q \le \chi^2_{\alpha/2}) = 1-\alpha$$

$$(2\text{-}53)$$

$$\Pr\left(\frac{(n-2)S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-2)S^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

حيث إن $\chi^2_{1-\alpha/2}$ و $\chi^2_{\alpha/2}$ قيمتان يتم الحصول عليهما من جدول تربيع كاي بدرجات حرية (n-2).

مثال:

باستخدام مثال نموذج انحدار وزن الطفل على عمره، اختبر دلالة معاملي النموذج عند مستوى دلالة إحصائية ٥% وحساب فترة ثقة ٩٥% للمعلمتين β_1 و β_0 وللمعلمة σ^2.

الحل:

أولًا: اختبار المعنوية:

اختبار معنوية β_1:

لإجراء اختبار المعنوية يتم أولًا صياغة فرض العدم. ولقياس أثر المتغير المستقل على المتغير التابع، يتم عادة اختبار ما إذا كانت قيمة المعلمة تتساوى بالصفر أم تختلف عنه اختلافًا ذا دلالة إحصائية. ولذلك في هذا المثال يتم اختبار الفرضية التالية:

$$H_0: \beta_1 = 0 \quad \textbf{مقابل} \quad H_1: \beta_1 \ne 0$$

وباستخدام المعادلة (2.47) يمكن حساب الإحصائية T_1 كما يلي:

$$T_1 = \frac{\hat{\beta}_1 - 0}{s.e(\beta_1)} = \frac{2.532 - 0}{0.1037} = 24.424$$

ولاختبار الذيلين بمستوى معنوية ٥% يتم حساب قيمة t النظرية عند درجات حرية (48) كما يلي:

$$t_{(n-2),\alpha/2} = t_{48,0.025} = 2.0106$$

وحيث إن قيمة T_1 المحسوبة أكبر من القيمة النظرية فإننا نرفض فرض العدم الذي ينص على أن قيمة المعلمة مساوية للصفر وبالتالي نقبل الفرض البديل ونقرر أن β_1 دال إحصائياً عند مستوى ٥%. وهذا يعني أن متغير العمر يسهم في تفسير تباين وزن الطفل.

اختبار معنوية β_0:

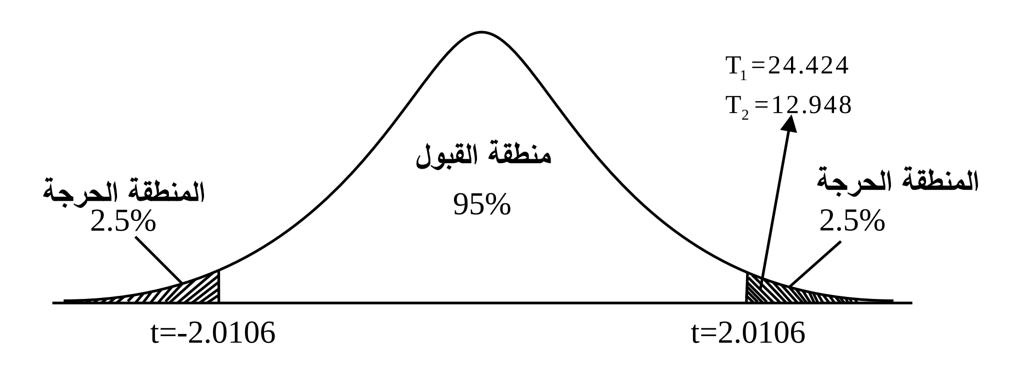
الفرض المراد اختباره هو:

$$H_0 : \beta_0 = 0 \quad \text{مقابل} \quad H_1 : \beta_0 \neq 0$$

وباستخدام المعادلة (2.48) يمكن حساب الإحصائية T_2 كما يلي:

$$T_2 = \frac{\hat{\beta}_0 - 0}{s.e(\beta_0)} = \frac{4.210 - 0}{0.32514} = 12.948$$

وحيث إن قيمة T_2 المحسوبة أكبر من القيمة النظرية فإننا نرفض فرض العدم الذي ينص على أن قيمة المعلمة مساوية للصفر، وبالتالي نقبل الفرض البديل ونقرر أن β_0 معنوية إحصائياً عند مستوى ٥%. ويوضح الشكل رقم (٢-٨) أن قيمتي إحصاء T_1 و T_2 المحسوبتين لمعاملي الانحدار تقع في منطقة الحرجة.



شكل رقم (٢-٨): منطقة القبول والرفض عند مستوى معنوية ٥% ودرجات حرية (٤٨).

كما يمكن إجراء اختبار ان تكون الزيادة السنوية في وزن الطفل خلال هذه الفترة من العمر أكبر من ٢ كيلوجرام، وعلى ذلك يمكن صياغة فرض ذي ذيل واحد على النحو التالي:

$$H_1: \beta > 2 \quad \text{مقابل} \quad H_0: \beta < 2$$

ولإجراء هذا الاختبار تستخدم المعادلة (2.47) التالية:

$$T = \frac{\hat{\beta}_1 - \beta_1^*}{s.e(\hat{\beta}_1)} = \frac{(2.532405 - 2)}{0.103685} = 5.1348$$

ونظراَ لأن قيمة T المشاهدة أكبر من قيمة t الجدولية ($t_{48,0.05} = 1.677$) فإننا نرفض فرض العدم القائل بان وزن الطفل أقل من ٢ وقبول الفرض البديل الذي ينص على أن الزيادة في وزن الطفل سنوياً أكبر من ٢ كيلوجرام في المتوسط.

ثانيًا: فترات الثقة:

* فترة الثقة للمعلمة β_1:

بالتعويض في المعادلة (2.51) نجد أن فترة الثقة للمعلمة عند مستوى معنوية (٥%) هي:

$$\hat{\beta}_1 \pm t_{(n-2),\alpha/2} s.e(\hat{\beta}_1) = 2.532 \pm t_{48,0.025} \times 0.1037 = 2.532 \pm 2.0106 \times 0.1037$$

إذن الفترة (٢،٣٢٤،٢،٧٤١) هي فترة ثقة بدرجة ٩٥% للمعلمة β_1، أي أننا على درجة ثقة بنسبة ٩٥% من أن قيمة المعلمة الحقيقية تقع في المدى من ٢،٣٢٤ إلى ٢،٧٤١. حيث يلاحظ ضيق فترة الثقة مما يعكس ثبات قيمة المقدر.

*فترة الثقة للمعلمة β_0:

بالتعويض في المعادلة (2.52) نجد أن فترة الثقة للمعامل الثابت عند مستوى معنوية (5%) هي:

$$\hat{\beta}_0 \pm t_{(n-2),\alpha/2} s.e(\hat{\beta}_0) = 4.210 \pm 2.0106 \times 0.325$$

إذن الفترة (٣،٥٥٦، ٤،٨٦٤) هي فترة ثقة بنسبة ٩٥%؛ أي أننا على ثقة بدرجة ٩٥% من أن قيمة المعلمة تقع في هذه الفترة. ويلاحظ أيضًا ضيق فترة الثقة مما يعكس ثبات قيمة المُقدر.

*فترة الثقة للمعلمة σ^2:

لحساب فترة الثقة يتم أولًا إيجاد القيم التالية:

$$S^2 = 2.3245, \qquad \chi^2_{48;0.025} = 69.023, \qquad \chi^2_{48;0.975} = 30.755$$

ويتم حساب فترة الثقة بالتعويض في الصيغة (2.53) كما يلي:

$$\Pr(\frac{(n-2)S^2}{\chi^2_{48;0.025}} < \sigma^2 < \frac{(n-2)S^2}{\chi^2_{48;0.975}} = \Pr(\frac{48x2.3245}{69.023} < \sigma^2 < \frac{48x2.3245}{30.755}) = \Pr\left(1.616 < \sigma^2 < 3.628\right) = 0.95$$

إذن فترة ثقة ٩٥% للمعلمة σ^2 هي ١,٦٢ و٣,٦٣

٢-١٠ جودة التوفيق (Goodness of Fit):

في هذا الجزء من الفصل سيتم الإجابة عن السؤال التالي: هل التغير/التباين في قيم المتغير التابع يرجع إلى التغير في قيم المتغير المستقل؟ أم يرجع إلى عوامل عشوائية أو إلى متغيرات أخرى؟ أو بمعنى آخر ما قدرة نموذج الانحدار في تفسير تباين/تغير المتغير التابع؟

٢-١٠-١ التغير/التباين المفسر والتغير/التباين غير المفسر (Variation Explained and Unexplained):

يقاس تباين المتغير التابع (Y) بمجموع مربعات انحرافات قيم Y عن الوسط الحسابي ($\bar{y}$). ويمكن تجزئة تباين المتغير التابع إلى جزأين: جزء تم تفسيره باستخدام نموذج الانحدار والجزء الآخر لم يتم تفسيره. ويتضح من الشكل رقم (٢-٩) أنه إذا رسمنا خط الانحدار المقدر وحددنا قيمة معينة للمتغير التابع (Y_i) فإن قيمة المتغير التابع المقدرة ($\hat{y}_i$) المناظرة تكون قد تحددت تماماً ويكون الفرق بين ($\hat{y}_i$) والوسط الحسابي ($\bar{y}$) هو فرق يعتمد على قيمة المتغير المستقل (X_i) . ويمثل هذا الجزء ($\hat{y}_i$-$\bar{y}$) التباين أو التغير في قيم Y_i الذي يعزى إلى التباين أو التغير في قيم X، أي التغير الذي حدث في Y نتيجة التغير في X. أما الجزء الثاني وهو (y_i-$\hat{y}_i$) فيمثل التباين الذي تبقى بعد طرح التباين المفسر، ويعتمد على عوامل غير معروفة. ورياضياً يمكن كتابة انحراف القيمة رقم i للمتغير التابع عن متوسط العينة كما يلي:

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + y_i - \hat{y}_i \qquad (2\text{-}54)$$

الانحراف الكلي = الانحراف المفسر + الانحراف غير المفسر

وبتربيع وجمع طرفي المعادلة (2.52) نحصل على:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (2.55)$$

Total Sum of Squares (TSS) = Explained Sum of Squares (ESS) + Residual Sum of Squares (RSS)

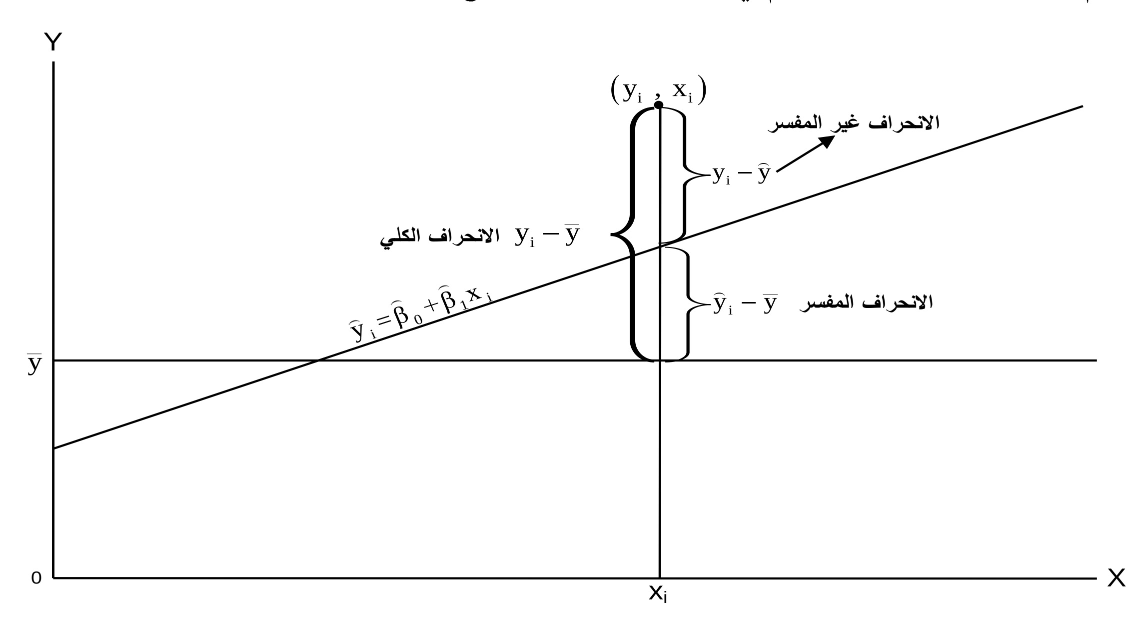
مجموع المربعات الكلي = مجموع مربعات الانحدار (مجموع المربعات التي تم تفسيرها) + مجموع مربعات البواقي (لم يتم تفسيرها)

وبما أن $\hat{y}_i - \bar{y} = \hat{\beta}_1(x_i - \bar{x})$ فإنه يمكن إعادة كتابة المعادلة (2.55) كما يلي:

$$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \hat{\beta}_1^2 \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 + \sum_{i=1}^{n} e_i^2 \qquad (2\text{-}56)$$

وتعـرف المعادلـة (2.55) أو(2.56) بمعادلـة تحليـل الانحـدار الأساسـية (Fundamental Equation of Regression Analysis). ويلاحظ من هاتين المعادلتين إذا كان مجموع مربعات البواقي مساوياً للصفر فيعني ذلـك أن توفيـق المعادلـة قـد تـم بصورة تامة؛ أي أن القيمة المشاهدة مساوية تماماً للقيمة المتنبأ بها لكل مشاهدة، وبذلك تقع كل القيم المشاهدة على خط الانحدار. وكلما كان مجموع مربعات البواقي كبيراً قل التباين الذي يفسره نموذج الانحدار وبالتالي ضعف أو عـدم قدرة النموذج في التنبؤ بقيم المتغير التابع. وهناك ثلاثـة عوامـل تسـهم في تضـخم قيمـة مجمـوع مربعـات البـواقي – وبالتالي تدنية قيمة مجموع المربعات التي تم تفسيرها - هي:

- وجود تباين كبير في قيم المتغير التابع.
- عدم تناسب فرضية العلاقة الخطية.
- عدم إدخال متغير/ متغيرات تسهم في تفسير تباين المتغير التابع.



شكل رقم (٢-٩): مكونات التباين/التغير الكلي للمتغير التابع

٢-١٠-٢ معامل التحديد (Coefficient of Determination):

بقسمة طرفي المعادلة (2.55) على مجموع المربعات الكلي (TSS) نحصل على:

$$\frac{TSS}{TSS} = \frac{ESS}{TSS} + \frac{RSS}{TSS} \qquad (2\text{-}57)$$

ويمكننا الآن تعريف معامل التحديد (r^2) كما يلي:

$$r^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS} \quad (2\text{-}58)$$

ويمكن كتابة معادلة معامل التحديد أعلاه حسب الصيغة التالية:

$$r^2 = \frac{\hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \hat{\beta}_1^2 \left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \right) = \frac{\left(\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2\text{-}59)$$

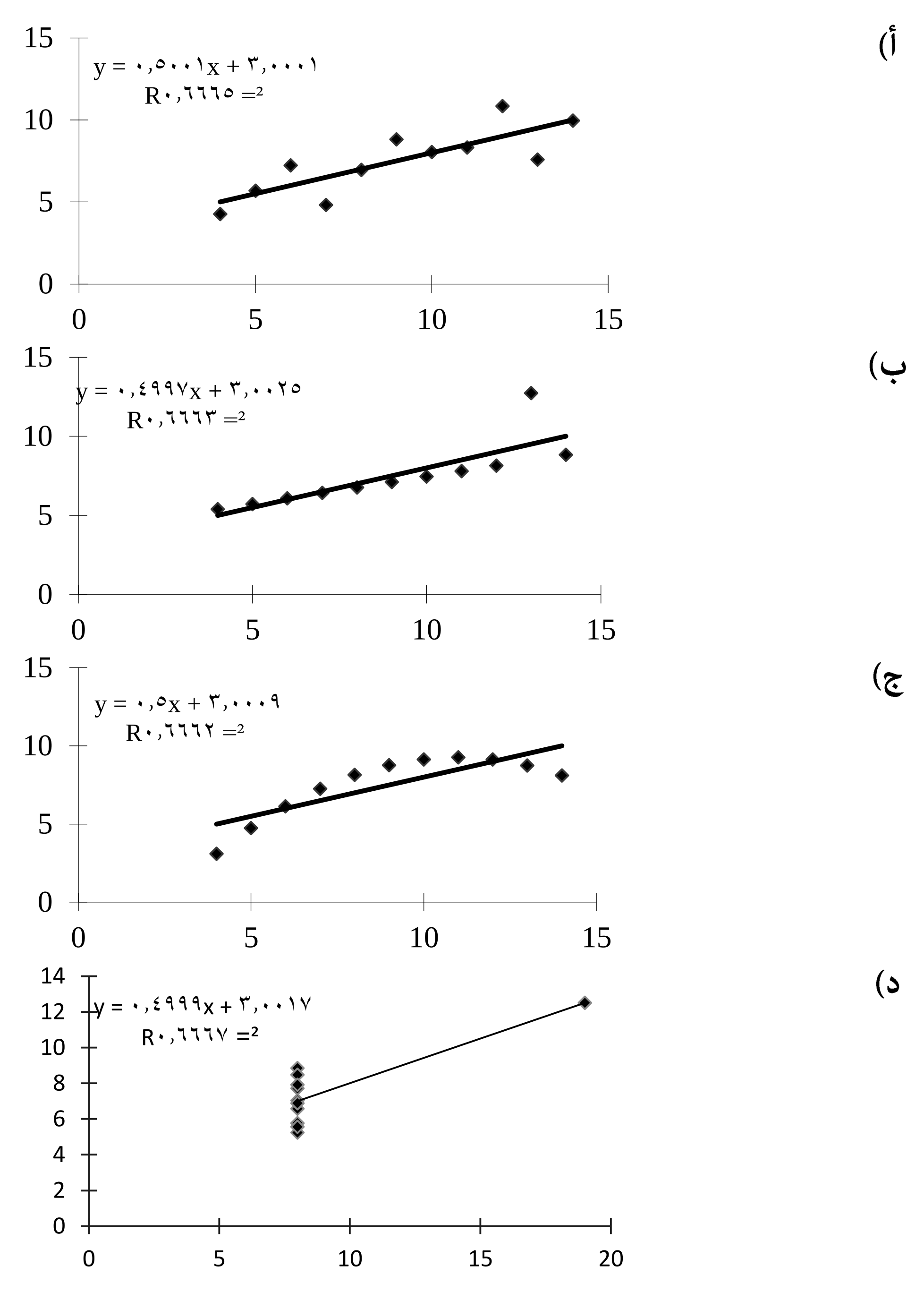
ويعتبر معامل التحديد – نسبة التباين/التغير المفسر لإجمالي التباين- من أكثر المقاييس استخداماً لقياس جودة التوفيق. ويقيس هذا المعامل نسبة التباين التي يفسرها نموذج الانحدار لإجمالي التباين في قيم المتغير التابع Y. ومن خصائص معامل التحديد نورد ما يلي:

- يأخذ معامل التحديد قيماً غير سالبة.
- تُراوح قيم معامل التحديد ما بين الصفر والواحد الصحيح، أي:

$$0 \le r^2 \le 1$$

عندما تكون قيمة معامل التحديد مساوية للواحد الصحيح يعني أن هناك علاقة تامة بين المتغيرين التابع والمستقل وإذا كانت القيمة مساوية للصفر فإن ذلك يعني عدم وجود علاقة بين المتغيرين.

- إن معامل التحديد لا يقيس دائماً مدى ملاءمة نموذج الخط المستقيم. ولتوضيح هذه الخاصية قام انسكومب (Anscombe 1973) بتحليل أربع مجموعات من البيانات موضحة بالأشكال رقم (٢-١٠-أ)، (٢-١٠-ب)، (٢-١٠-ج) و (٢-١٠-د). وعلى الرغم من أن قيم معامل التحديد متقاربة($r^2 \approx 0.67$) إلا أن أشكال الانتشار الممثلة لهذه المجموعات من البيانات مختلفة تماماً. فالشكل (٢-١٠-أ) يوضح أن هناك علاقة خطية بين المتغير التابع (Y) والمتغير المستقل (X). في حين يتضح من الشكل (٢-١٠-ب) أن هناك علاقة خطية بين المتغيرين إلا أنه توجد حالة واحدة شاذة (Outlier) ربما تكون مؤثرة على مقدرات المربعات الصغرى. ويوضح الشكل (٢-١٠-ج) أن العلاقة بين المتغير التابع والمستقل غير خطية بل هي دالة تربيعية وإضافة الحد $\left(X^2\right)$ للنموذج ستزيد من قيمة معامل التحديد. أما الشكل الرابع (٢-١٠-د) فيوضح أن البيانات غير عادية (Unusual)، إذ إن جميع قيم المتغير المستقل متساوية ما عدا قيمة واحدة شاذة ومؤثرة في نفس الوقت على مقدرات المربعات الصغرى. وبما أن المتغير المستقل يضم قيمتين فقط فإنه من الصعوبة معرفة شكل العلاقة خطية أم لا؟. وتؤكد هذه الأشكال ضرورة البدء برسم شكل الانتشار قبل الشروع في بناء نموذج الانحدار.



شكل رقم (٢-١٠) أشكال انتشار لأربع مجموعات من البيانات لها معاملات تحديد متماثلة

المصدر: Anscombe (1973) pp.17-22

٢-١٠-٣ الارتباط الخطي البسيط:

بأخذ الجذر التربيعي لمعامل التحديد (r^2) نحصل على مقياس يعرف بمعامل الارتباط الخطي البسيط (Simple Linear Correlation) ويسمى أيضا بمعامل ارتباط بيرسون (Pearson). وبأخذ الجذر التربيعي للمعادلة (2.59) نحصل على معامل الارتباط الخطي البسيط كما يلي:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \hat{\beta}_1 \frac{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (2\text{-}60)$$

ويستخدم معامل الارتباط لقياس قوة واتجاه العلاقة الخطية بين متغيرين عشوائيين يتبعان التوزيع الطبيعي وليس مهمًا أيهما المتغير التابع أو المتغير المستقل. وفيما يلي أهم خصائص معامل الارتباط البسيط.

١) تُراوح قيم معامل الارتباط بين سالب واحد وموجب واحد ($-1 \le r \le +1$). إذا كان معامل الارتباط موجبًا يعني ذلك أن العلاقة بين المتغيرين طردية، أي أن الزيادة في قيم المتغير الأول تصاحبها زيادة في قيم المتغير الآخر وكذلك النقص في قيم المتغير الأول يصاحبه أيضًا نقص في قيم المتغير الثاني. وأما إذا كان معامل الارتباط سالبًا فيعني ذلك أن العلاقة عكسية بين المتغيرين، أي أن الزيادة في المتغير الأول يقابلها نقص في قيم المتغير الثاني والعكس صحيح. وتعتبر العلاقة بين المتغيرين قوية كلما اقتربت القيمة المطلقة للمعامل من الواحد الصحيح، وتوصف العلاقة بالضعيفة كلما اقتربت قيمة المعامل المطلقة للصفر. ويساعد رسم الانتشار في تحديد اتجاه العلاقة بين المتغيرين. فمثلًا يوضح الشكل رقم (٢-١١-أ) أن هناك علاقة طردية بين المتغيرين Y وX، ويشير الشكل (٢-١١-ب) إلى وجود علاقة عكسية بين المتغيرين، وأما إذا كان الشكل أشبه بمنحنى (شكل رقم ٢-١١-ج) فإن العلاقة بين المتغيرين تسمى علاقة غير خطية. ويشير الشكل رقم (٢-١١-د) إلى عدم وجود علاقة ارتباط خطية بين المتغيرين.

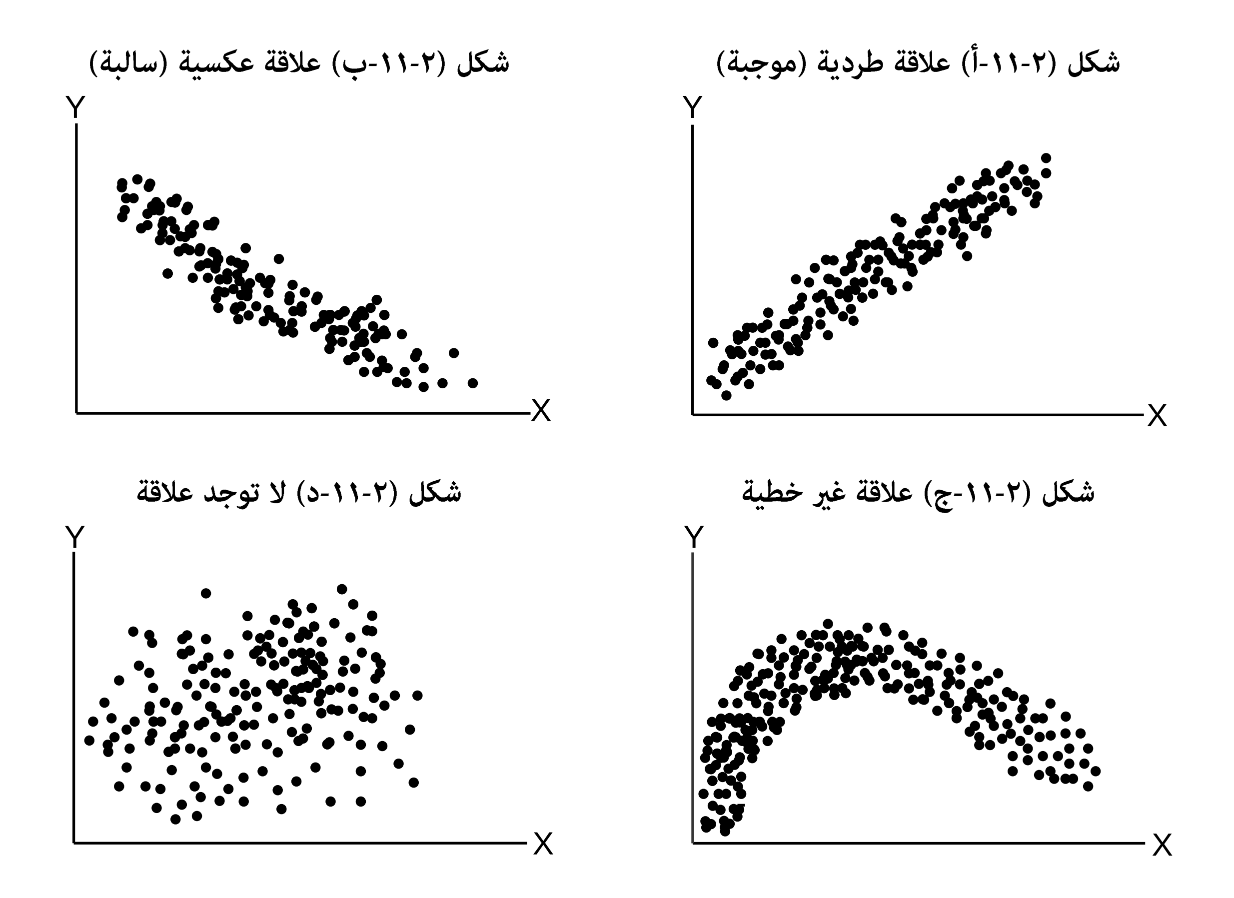
٢) قيمة معامل الارتباط لا تعتمد على وحدات قياس المتغيرين. فمثلًا معامل ارتباط الطول مقاسًا بالبوصات مع الوزن مقاسًا بالأرطال له نفس قيمة معامل ارتباط الطول مقاسًا بالأمتار مع الوزن مقاسًا بالكيلوجرامات.

٣) إشارة معامل الارتباط هي نفس إشارة معامل الانحدار ($\hat{\beta}_1$)، فإذا كان معامل الارتباط موجبًا يعني أن ميل خط الانحدار موجبًا أيضًا وإذا كان معامل الانحدار سالبًا يعني أن ميل خط الانحدار سالبًا.

٤) إن وجود علاقة ارتباط بين متغيرين لا يعني وجود علاقة سببية.

٥) يمكن حساب معامل التحديد (r^2) بتربيع معامل الارتباط البسيط (r) .

٦) بما أن معامل الارتباط يتم حسابه من مشاهدات العينة فيعتبر المعامل مقدر لمعلمة المجتمع المجهولة التي تعرف بمعامل ارتباط المجتمع ويرمز له بـ ρ



اختبار معنوية معامل الارتباط (r):

لاختبار فرض العدم القائل بأن معامل ارتباط المجتمع ρ يساوي الصفر في مقابل الفرض البديل القائل بأنه يختلف معنوياً عن الصفر($H_0:\rho=0$ مقابل $H_1:\rho\neq 0$)، تستخدم إحصائية T حيث

$$T=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2} \qquad (2\text{-}61)$$

لها توزيع t بدرجات حرية (n-2). ويلاحظ أن اختبار معنوية معامل الارتباط يكافئ اختبار معنوية ميل الانحدار. وهذا يعني إذا كان $\hat{\beta}_1$ دالاً إحصائياً فإن r أيضًا دالٌ إحصائياً والعكس صحيح وذلك لأن كليهما يقيس وجود أو عدم وجود علاقة خطية بين المتغيرين Y وX.

٢-١٠-٤ جدول تحليل التباين (Analysis of Variance Table):

يتم عادة وضع مكونات مجموع المربعات الكلي ودرجات الحرية المصاحبة في جدول يعرف **بجدول تحليل التبـاين**. وترجع هذه التسمية إلى أن المعلومات الأساسية التي يحتويها الجدول تشمل تقدير التبـاين. ويسـتخدم هـذا الجـدول للإجابة عن الأسئلة الاستدلالية التالية:

- هل تختلف قيمة الميل الحقيقي عن الصفر؟
- ما قوة العلاقة الخطية بين المتغيرين التابع والمستقل؟
- هل النموذج الخطي مناسبٌ لتوفيق بيانات المتغيرين التابع والمستقل؟

إن الإجابة عن هذه الأسئلة تتطلب تحليل تباين قيم المتغـير التـابع Y. وكـما سـبق ذكـره، فـإن مجمـوع المربعـات للمتغير التابع Y يساوي مجموع مربعات الانحدار زائدًا مجموع مربعات البواقي، أي:

$$\text{TSS} \qquad = \text{ESS} \qquad + \text{RSS}$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\hat{\beta}_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}e_i^{\,2}$$

ويصاحب كل مجموع مربعات درجات حرية (Degrees of freedom) أو اختصاراً (DF) وتسـتخدم درجـات الحريـة في الإحصـاء للتعبـير عـن المشـاهدات أو المعـالم المسـتقلة خطيًـا في مجمـوع المربعـات. فمـثلًا القـيم $(Y_1,\ Y_2,\ \ldots,Y_n)$ هـي n مـن المشـاهدات المسـتقلة. وبمـا أن $(\sum_{i=1}^{n}(y_i-\bar{y})=0)$ فـإن المقـادير $(y_i-\bar{y})$ لا تكـون مستقلة خطيًا لأنه يمكن اشتقاق واحد منهما من الآخرين ولذلك يكون لمجموع مربعاتها $\sum_{i=1}^{n}(y_i-\bar{y})^2$ درجات حرية (n-1). وكذلك أن درجات الحرية المقابلة لمجموع مربعات البـواقي هـي (n-2) لفقـدنا درجتـي حريـة بسـبب تقـدير معلمتـي نمـوذج الانحـدار البسـيط مـن أجـل الحصـول عـلى القيمـة المقـدرة $(\hat{y}_i)$. أمـا مجمـوع مربعـات الانحـدار $\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ فيقابله درجة حرية واحدة. ويرجع ذلك لتقدير معلمتـي نمـوذج الانحـدار وعـدم اسـتقلالية المقـادير $(\hat{y}_i-\bar{y})$ لأن مجموعها يساوي صفرًا. وبصورة عامة؛ تساوي درجات الحرية المصاحبة لمجموع مربعات الانحدار عدد معالم نموذج الانحدار المقدرة ناقصًا واحد.

متوسط مربعات البواقي والانحدار:

متوسط المربعات هو خارج قسمة مجموع المربعات على درجات الحرية. ومن هذا التعريف يمكن كتابة متوسطي مربعات الانحدار والبواقي على النحو التالي:

متوسط مربعات الانحدار = $\frac{\text{مجموع مربعات الانحدار}}{\text{عدد درجات الحرية}}$ = $\frac{\text{مجموع مربعات الانحدار}}{1}$ = مجموع مربعات الانحدار

$$MESS=\frac{ESS}{1}=\hat{\beta}_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

متوسط مربعات الانحدار = $\frac{\text{مجموع مربعات البواقي}}{\text{عدد درجات الحرية}}$

$$MRSS=\frac{RSS}{n-2}=\frac{\sum_{i=1}^{n}e_i^2}{n-2}=S^2$$

القيمة المتوقعة لمتوسطات المربعات:

يعتبر التعرف على القيمة المتوقعة لمتوسطات المربعات خطوة أساسية لاختبار الفروض بطريقة تحليل التباين. وعلى ذلك فإن الخطوة التالية هي إيجاد التوقع لمتوسطات مربعات الانحدار والبواقي كما يلي:

- القيمة المتوقعة لمتوسط مربعات البواقي هي:

$$E(MRSS)=E\left(\frac{\sum_{i=1}^{n}e_i^2}{n-2}\right)=E(S^2)=\sigma^2$$

- القيمة المتوقعة لمتوسط مربعات الانحدار هي:

$$E(ESS)=E(\hat{\beta}_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2)=\beta_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+\sigma^2$$

ولإثبات ذلك نستخدم النتيجة التالية:

$$var\left(\hat{\beta}_1\right)=E\left(\hat{\beta}_1^2\right)-\left[E\left(\hat{\beta}_1\right)\right]^2$$

$$\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=E\left(\hat{\beta}_1^2\right)-\beta_1^2 \Rightarrow E\left(\hat{\beta}_1^2\right)=\beta_1^2+\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

ومن ثم فإن

$$E(ESS)=E\left(\hat{\beta}_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)=\left(\beta_1^2+\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right)\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\beta_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+\sigma^2$$

ويلاحظ ما يأتي من القيم المتوسطة لمتوسطات المربعات:

- القيمة المتوقعة لمتوسط مربعات البواقي هي σ^2 بصرف النظر عن وجود أو عدم وجود علاقة خطية بين المتغيرين Y وX، أي أن القيمة المتوقعة تساوي σ^2 إذا كانت قيمة المعلمة β_1 مساوية للصفر أو تختلف عنه.
- القيمة المتوقعة لمربعات الانحدار تساوي σ^2 إذا كانت قيمة المعلمة β_1 مساوية للصفر. أما إذا كانت قيمة المعلمة تختلف عن الصفر $(\beta_1 \neq 0)$، فإن القيمة المتوقعة لمتوسط مربعات الانحدار تكون أكبر من ذلك لأن $\beta_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$ قيمة موجبة. ولذلك من الطبيعي أن نقارن بين متوسط مجموع مربعات الانحدار (MESS) و متوسط مجموع مربعات البواقي (MRSS) عند اختبار معنوية ميل نموذج الانحدار.

اختبار F:

بعد تجزئة التغير/التباين الكلي في Y يبقى أن نختبر ما إذا كان الانحدار الخطي قد فسر جزءاً كبيراً ذات دلالة أم لا، أي أن نختبر ما إذا كان التباين/التغاير المفسر أكبر من التباين غير المفسر. والفرض المراد اختباره في نموذج الانحدار الخطي البسيط هو:

$$H_0 : \beta_1 = 0 \quad \text{مقابل} \quad H_1 : \beta_1 \neq 0$$

ولإجراء هذا الاختبار تستخدم إحصائية F_0 حيث

$$F_0 = \frac{MESS}{MRSS}$$

أي

$$F_0 = \frac{\text{متوسط مجموع مربعات الانحدار}}{\text{متوسط مجموع مربعات البواقي}} = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$

وتشير القيمة العالية لإحصائية F_0 إلى وجود علاقة معنوية بين المتغير التابع والمتغير المستقل، مؤدية إلى رفض فرض العدم.

توزيع F_0:

لصياغة قاعدة القرار الإحصائي لا بد من معرفة توزيع معاينة F_0. وعند فرض العدم $(H_0 : \beta_1 = 0)$ نجد أن كل مشاهدة (Y_i) لديها متوسط واحد $(\mu = \beta_0)$ وتباين واحد أيضًا مقداره σ^2 مع ملاحظة أن المتغيرين

$$\frac{RSS}{\sigma^2} \text{ و } \frac{ESS}{\sigma^2}$$

متغيران مستقلان ولهما توزيع مربع كاي. والآن تأخذ إحصائية الاختبار F_0 الصيغة التالية:

$$F_0 = \frac{ESS/\sigma^2}{1} \div \frac{RSS/\sigma^2}{n-2} = \frac{MESS}{MRSS}$$

وطبقًا للنظرية الإحصائية فإن إحصائية F_0 حيث

$$F_0 = \frac{\chi^2_1}{1} \div \frac{\chi^2_{(n-2)}}{n-2} \sim F_{1,n-2}$$

تتبع توزيع F بدرجات حرية ١ و$(n-2)$. ويمكن إعادة كتابة إحصائية F_0 كما يلي:

$$F_0 = \frac{\hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{S^2} \sim F_{1,n-2} \qquad (2\text{-}62)$$

وتوضع عادة مجموعات المربعات ودرجات الحرية المصاحبة لها وإحصائية F_0 في جدول تحليل التباين التالي:

جدول تحليل التباين لنموذج الانحدار الخطي البسيط

مصدر التغير/التباين	درجات الحرية	مجموع المربعات	متوسط المربعات	قيمة F_0
الانحدار	1	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$	$\hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$	$\frac{\hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{S^2}$
البواقي	n-2	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$	$\frac{\sum_{i=1}^{n} e_i^2}{(n-2)} = S^2$	
المجموع	n-1	$\sum_{i=1}^{n}(y_i - \bar{y})^2$		

وبعد تحديد مستوى المعنوية (α) يمكننا تحديد قاعدة القرار التالية:

- إذا كانت قيمة F_0 أكبر من قيمة $F_{1,n-2}$ الجدولية نرفض فرض العدم $(H_0 : \beta_1 = 0)$ وبالتالي قبول الفرض البـديل $(H_0 : \beta_1 \neq 0)$، أي أن قيمة المعلمة β_1 تختلف عن الصفر ونحكم بوجود علاقة خطية بين المتغيرين Y وX.
- أما إذا كانت قيمة F_0 أقل من قيمة $F_{1,n-2}$ فنقول ليس لدينا دليل كاف لرفض فرض العدم وبالتالي نحكم بعدم وجود علاقة خطية بين المتغيرين Y وX ذات دلالة إحصائية، لأن معنى $\beta_1 = 0$ أن تكون قيمة $\hat{y} = \hat{\beta}_0$، أي Y تكون قيمة ثابتة ولا يكون لقيم X أي أثر خطي على Y.

ملاحظات:

- ويلاحظ أن اختبار F_0 يكافئ اختبار (t) لاختبار وجود علاقة خطية بين المتغيرين Y وX. ويمكن اشتقاق أي منهما من الآخر إذ إن قيمة F_0 تساوي مربع قيمة (t) كما يتضح ذلك من إعادة كتابة إحصائيتي الاختبار لكل منهما كما يلي:

$$T=\frac{\beta_1}{s.e(\beta_1)}=\frac{\beta_1\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}{S} \quad\Rightarrow\quad T^2=\frac{\beta_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2}{S^2}=F_0$$

-يمكن إجراء اختبار F_0 بمعلومية معامل التحديد r^2 . ولتوضيح ذلك يمكن كتابة إحصائية F_0 كما يلي:

$$F_0=\frac{ESS}{RSS/(n-2)}$$

وبقسمة بسط ومقام هذه المعادلة على مجموع المربعات الكلي (TSS) نحصل على:

$$F_0=\frac{ESS/TSS}{(RSS/TSS)/(n-2)}=\frac{ESS/TSS}{(1-ESS/TSS)/(n-2)}$$

وبما أن $r^2=\frac{ESS}{TSS}$ فإن:

$$F_0=\frac{r^2}{(1-r^2)/(n-2)} \sim F_{1,\,n-2} \qquad (2-63)$$

- إن الصيغة الإحصائية F_0 الواردة أعلاه هي صيغة عامة لاختبار قدرة نموذج الانحدار في تفسير تباين/تغير المتغير التابع مهما كان عدد المتغيرات المستقلة في نموذج الانحدار مع ملاحظة اختلاف درجات الحرية التي تتغير تبعاً لعدد المتغيرات المستقلة.

مثال:

من بيانات مثال نموذج انحدار وزن الطفل على عمره، احسب معامل التحديد ومعامل الارتباط البسيط، وجدول تحليل التباين.

الحل:

معامل التحديد r^2:

الجدول رقم (٢-٧) يعطي الحسابات المطلوبة لحساب معامل التحديد، وهي:

$$\sum_{i=1}^{50}(x_i-\bar{x})^2=216.218408,\qquad \sum_{i=1}^{50}(y_i-\bar{y})^2=1498.199,\qquad \hat{\beta}_1=2.532$$

$$\sum_{i=1}^{50}(x_i-\bar{x})(y_i-\bar{y})=547.5526,\qquad \sum_{i=1}^{50}(\hat{y}_i-\bar{y})^2=1386.6248$$

وبالتعويض في الصيغة (2.58) أو(2.59) نتحصل على:

$$r^2=\frac{\sum_{i=1}^{50}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{50}(y_i-\bar{y})^2}=\frac{1386.6248}{1498.199}=0.926$$

أو

$$r^2=\frac{\hat{\beta}_1^2\sum_{i=}^{50}(x_i-\bar{x})^2}{\sum_{i=1}^{50}(y_i-\bar{y})^2}=\frac{(2.532)^2\times 216.218408}{1498.199}=0.926$$

هذا يعني أن حوالي ٩٣% من التغير/التباين في أوزان الأطفال يمكن تفسيره بالتغير في أعمارهم. وتشير هذه النتيجة إلى أن نموذج وزن الطفل المقدر يتمتع بمقدرة تفسيرية عالية.

معامل الارتباط الخطي البسيط (r):

بالتعويض في الصيغة (2.60) نتحصل على:

$$r=\frac{\sum_{i=1}^{50}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{50}(x_i-\bar{x})^2\sum_{i=1}^{50}(y_i-\bar{y})^2}}=\frac{547.55256}{\sqrt{216.218408\times 1498.199}}=0.962$$

أو

$$r=\hat{\beta}_1\frac{\sqrt{\sum_{i=1}^{50}(x_i-\bar{x})^2}}{\sqrt{\sum_{i=1}^{50}(y_i-\bar{y})^2}}=2.532\times\frac{\sqrt{216.218408}}{\sqrt{1498.199}}=0.962$$

وهذا يعني أن الارتباط بين وزن الطفل وعمره ارتباط طردي وقوي، أي كلما كبر الطفل زاد وزنه.

تحليل التباين:

من الجدول رقم (٢-٥) نستطيع حساب القيم اللازمة لاستخدام صيغة إحصائية F_0:

- مجموع المربعات الكلي (TSS):

$$TSS = \sum_{i=1}^{50}(y_i - \bar{y})^2 = 1498.199$$

- مجموع مربعات البواقي (RSS):

$$RSS = \sum_{i=1}^{50}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{50} e_i^2 = 111.5744$$

- مجموع مربعات الانحدار (ESS):

$$ESS = TSS - RSS = 1498.199 - 111.5744 = 1386.6248$$

أو

$$ESS = \hat{\beta}_1^2 \sum_{i=1}^{50}(x_i - \bar{x})^2 = (2.532)^2 \times 216.218408 = 1386.6248$$

ويتم حساب درجات الحرية المصاحبة لمجاميع المربعات كما يلي:

درجات الحرية لمجموع مربعات الانحدار =عدد معالم النموذج – ١=٢-١=١

درجات الحرية لمجموع مربعات البواقي =عدد المشاهدات – عدد المعالم = ٥٠-٢=٤٨

درجات الحرية لمجموع المربعات الكلي =عدد المشاهدات – ١= ٥٠-١=٤٩

= **درجات الحرية مجموع مربعات البواقي + درجات الحرية مجموع مربعات الانحدار**

= ٤٨ + ١=٤٩

ويتم حساب متوسطي مربعات الانحدار والبواقي كما يلي:

متوسط مربعات الانحدار = $\frac{\text{مجموع مربعات الانحدار}}{\text{درجات الحرية}} = \frac{1386.6248}{1} = 1386.6248$

متوسط مربعات البواقي = $\frac{\text{مجموع مربعات البواقي}}{\text{درجات الحرية}} = \frac{111.5744}{48} = 2.324467$

ومن ثم يتم إحصائية F_0 كما يلي:

$$F_0 = \frac{MESS}{MRSS} = \frac{1386.6248}{2.324467} = 596.5346$$

أو

$$F_0 = \frac{\beta_1^2 \sum_{i=1}^{50}(x_i-\bar{x})^2}{S^2} = \frac{(2.53240492)^2 \times 216.218408}{2.324467} = 596.5346$$

ويتم وضع هذه القيم في جدول تحليل التباين التقليدي أدناه:

مصدر التغير/التباين	درجات الحرية	مجموع المربعات	متوسط المربعات	قيمة إحصائية F_0
الانحدار (التباين المفسر)	١	١٣٨٦,٦٢٤٨	١٣٨٦,٦٢٤٨	٥٩٦,٥٣٤٦
البواقي (التباين غير المفسر)	٤٨	١١,٥٧٤٤	٢,٣٢٤٤٦٧	
المجموع (التباين/التغير الكلي)	٤٩	١٤٩٨,١٩٩		

وحيث إن القيمة (F_0=596.5346) أكبر بكثير من القيمة الحرجة لتوزيع F عند درجتي حرية ١ و٤٨ ومستوى معنوية ١% ($F_{1,48,0.01}$=7.1942) مما يجعلنا نرفض فرض العدم عند مستوى عال من الدلالة ونحكم بوجود علاقة خطية بين وزن الطفل وعمره، أو بمعنى آخر أن متغير عمر الطفل يسهم بمستوى معنوي في تفسير تباين وزن الطفل.

جدول رقم (٢-٥): الحسابات اللازمة لتقدير معامل التحديد وحساب احصائية F_0

رقم المشاهدة	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	e_i^2	$(\bar{y}_i-\bar{y})$
1	1.811716	0.878669	0.426148	0.094344	2.73291792
2	34.17572	15.50827	7.037348	0.760321	45.1310372
3	13.35172	6.749669	3.412148	1.048286	21.8823589
4	46.86772	11.31507	2.731748	7.077947	17.51890289
5	2.735716	1.682449	1.034696	0.850014	6.635581687
6	1.833316	1.824109	1.814948	4.233948	11.63939603
7	140.3277	45.28489	14.6138	4.687755	93.71938964
8	8.099716	3.053189	1.1509	0.016702	7.380806616
9	5.503716	3.103289	1.7498	1.007745	11.22159704
10	28.57972	16.42719	9.4421	5.93202	60.55289135
11	0.427716	0.769889	1.3858	5.415613	8.887237857
12	28.57972	11.08119	4.2965	0.009376	27.55377431
13	0.427716	0.769889	1.3858	5.415613	8.887237857
14	18.88772	1.750569	0.162248	11.06193	1.040507514
15	78.25172	34.52417	15.23185	1.076344	97.68297762
16	1.331716	0.977669	0.717748	0.98298	4.602970495
17	14.79172	7.318169	3.620648	0.946068	23.21948497
18	0.119716	-0.12013	0.120548	1.50124	0.7730823
19	17.25572	8.005589	3.7141	0.527731	23.81879962
20	23.48372	15.66615	10.451	11.16067	67.02301674
21	8.099716	3.053189	1.1509	0.016702	7.380806616
22	117.6357	41.46209	14.6138	1.35751	93.71938964
23	3.407716	1.205069	0.426148	0.03719	2.73291792
24	53.96372	21.32397	8.426248	2.57E-05	54.03815662
25	21.65972	9.388049	4.069096	0.20645	26.09541548
26	23.56132	9.791489	4.069096	0.064703	26.09541548
27	13.35172	5.836169	2.551048	0.152691	16.36006029
28	11.19572	4.961449	2.198696	0.167322	14.10040061
29	31.96772	11.85757	4.398248	0.117677	28.20629183
30	28.57972	12.84537	5.773448	0.545918	37.02555212
31	40.27172	14.74049	5.3954	0.215045	34.60110207
32	0.715716	-0.50523	0.356648	5.561825	2.28720923
33	53.96372	21.32397	8.426248	2.57E-05	54.03815662

رقم المشاهدة	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	e_i^2	$(\hat{y}_i-\bar{y})$
34	19.32482	10.91439	6.164296	3.577602	39.53208953
35	0.023716	0.053469	0.120548	0.525989	0.7730823
36	37.87172	13.39849	4.7402	0.410174	30.39925554
37	44.27572	15.08595	5.140196	0.832714	32.96445976
38	4.639716	2.901869	1.814948	1.581698	11.63939603
39	4.639716	2.191049	1.034696	0.178052	6.635581687
40	14.79172	5.395169	1.967848	0.086167	12.61995514
41	70.62722	18.29719	4.7402	8.35469	30.39925554
42	48.35812	15.76611	5.140196	1.470233	32.96445976
43	21.19682	9.287189	4.069096	0.254386	26.09541548
44	54.81922	16.78635	5.140196	2.764011	32.96445976
45	77.51042	20.57671	5.462504	8.324744	35.03144503
46	21.65972	8.224549	3.122996	0.031946	20.02800553
47	31.96772	12.81875	5.140196	0.007651	32.96445976
48	47.66522	16.06699	5.41586	1.021287	34.73231358
49	46.97732	16.08771	5.509348	0.827989	35.3318591
50	76.63252	19.84707	5.140196	9.075346	32.96445976
المجموع	1498.199	547.5526	216.2184	111.5744	1386.624796

٢-١١ التقدير والتنبؤ (Estimation an Prediction):

بعد بناء نموذج الانحدار وفحصه والتأكد من استيفائه للفروض يصبح بالإمكان استخدامه للتقدير والتنبؤ. وتستخدم معادلة الانحدار المقدرة لتقدير القيمة المتوسطة للمتغير التابع المقابلة لقيمة معينة للمتغير المستقل (x_0) وللتنبؤ بقيمة مشاهدة جديدة للمتغير التابع المقابلة لقيمة مختارة للمتغير المستقل (x_0).

٢-١١-١ تقدير القيمة المتوسطة للمتغير التابع:

ينقسم تقدير القيمة المتوسطة للمتغير إلى نوعين هما: **تقدير النقطة** (Point Estimation) و**تقدير الفترة** (Interval Estimation).

مقدر النقطة:

يستخدم نموذج الانحدار المقدر التالي لتقدير القيمة المتوسطة ($\hat{Y}_0$):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

حيث يتم تعويض قيمة المتغير المستقل المراد عندها تقدير قيمة المتغير التابع (X_0) في الجانب الأيمن من المعادلة أعلاه للحصول على القيمة المقدرة كما يلي:

$$\hat{y}_0 = E(y/x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 \qquad (2\text{-}64)$$

حيث إن ($\hat{y}_0$) هو تقدير القيمة المتوسطة للمتغير التابع وهو مقدر غير متحيز للقيمة الحقيقية للمتغير التابع (y_0) عندما تكون قيمة المتغير المستقل مساوية لـ x_0، $\hat{\beta}_0$ و $\hat{\beta}_1$ معاملات نموذج الانحدار المقدرة باستخدام طريقة المربعات الصغرى.

فترة الثقة:

مما لا شك فيه أن القيمة المتوسطة للمتغير التابع المقدرة قد تختلف من القيمة الحقيقية؛ ويعرف الفرق بين القيمة المقدرة والقيمة الحقيقية للمتغير التابع بخطأ التقدير. ولقياس مدى دقة التقدير يتم عادة حساب الخطأ المعياري للقيمة المقدرة وإيجاد فترة الثقة. وفيما يلي خطوات تكوين فترة الثقة للقيمة المقدرة:

بما أن حد الخطأ (ε_i) يتبع التوزيع الطبيعي فإن القيمة المقدرة $\hat{y}_0$ تكون هي الأخرى تتبع التوزيع الطبيعي، أي:

$$\hat{y}_0 \sim N\left[\beta_0+\beta_1 x_0,\quad \sigma^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)\right] \qquad (2\text{-}65)$$

وأن القيمة المقدرة مقدر غير متحيز للقيمة الحقيقية (Y_0)

برهان:

لإثبات أن القيمة المتوسطة ($\hat{y}_0$) مقدر غير متحيز لـ y_0، يتم أخذ القيمة المتوقعة لطرفي المعادلة (2.64) كما يلي:

$$E(\hat{y}_0)=E(\hat{\beta}_0+\hat{\beta}_1 x_0)=\beta_0+\beta_1 x_0$$

وبذلك تصبح ($\hat{y}_0$) مقدراً غير متحيز لـ Y_0.

أما تباين ($\hat{y}_0$) فيمكن إيجاده كما يلي:

بأخذ تباين طرفي المعادلة (2.64) نحصل على:

$$\text{var}(\hat{y}_0)=\text{var}(\hat{\beta}_0)+x_o^2\,\text{var}(\hat{\beta}_1)+2x_o\,\text{cov}(\hat{\beta}_0,\hat{\beta}_1)$$

وبتعويض قيم تباين $\hat{\beta}_0$ و $\hat{\beta}_1$ وتغاير $\hat{\beta}_0$ و $\hat{\beta}_1$ في المعادلة أعلاه نحصل على:

$$\text{var}(\hat{y}_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}+x_0^2\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}-\frac{2x_0\bar{x}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

وبما أن $\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2$ فإن

$$\text{var}(\hat{y}_0)=\sigma^2\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2+nx_o^2-2nx_0\bar{x}}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)=\sigma^2\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n(x_0-\bar{x})^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)=\sigma^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

كما يمكن حساب التباين بطريقة أخرى كما يلي:

$$\operatorname{var}(\hat{y}_0) = \operatorname{var}(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \operatorname{var}(\bar{y} + \hat{\beta}_1 (x_0 - \bar{x}))$$
$$= \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{S_{xx}} = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$$

وبما أن قيمة (σ^2) مجهولة فإنه يستعاض عنها بالمقدر غير المتحيز (S^2)، وبالتالي يصبح مقدر تباين القيمة المتوسطة ($S^2(\hat{y}_0)$):

$$S^2(\hat{y}_0) = S^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$$

وبأخذ الجذر التربيعي لطرفي المعادلة نحصل على الخطأ المعياري للقيمة المتوسطة كما يلي:

$$s.e(\hat{y}_0) = S \sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)}$$

ولإيجاد فترة التنبؤ للقيمة المتنبأ بها يستخدم المتغير T حيث

$$T = \frac{\hat{y}_0 - E(y_0)}{s.e(\hat{y}_0)} \sim t_{n-2}$$

يتبع توزيع 't' بدرجات حرية (n-2). وبالطريقة المعتادة لتأسيس فترات الثقة كما سبق شرحه تكون فترة الثقة كما يلي:

$$\hat{y}_0 \pm t_{n-2,\alpha/2} s.e(\hat{y}_0) \qquad (2\text{-}66)$$

ويلاحظ أن مقدر التباين ($S^2(\hat{y}_0)$) يصل إلى أقل حد له عندما تكون قيمة (x_0) مساوية للوسط الحسابي ($\bar{x}$)، وبالتالي فإن فترة الثقة تعتمد على مدى قرب قيمة المتغير المستقل (x_0) المراد عندها التقدير بالقيمة المتوسطة للمتغير التابع للوسط الحسابي له.

٢-١١-٢ التنبؤ بمشاهدة جديدة أو التنبؤ الفردي:

التنبؤ بنقطة:

لا يوجد اختلاف في تنبؤ النقطة بين تقدير القيمة المتوسطة والتنبؤ الفردي للمتغير التابع، حيث تستخدم في الحالتين المعادلة (2.64) للتقدير أو التنبؤ. إلا أن فترة الثقة تختلف في كل حالة، إذ إنها أكبر في حالة التنبؤ الفردي من حالة التنبؤ بالقيمة المتوسطة؛ ذلك لكبر الخطأ المعياري للتنبؤ الفردي. وينظر للتنبؤ الفردي على أنه التنبؤ بمشاهدة جديدة خارج نطاق قيم مشاهدات العينة، حيث يفترض الباحث ثبات العلاقة بين المتغيرين التابع والمستقل إذا احتوت العينة على المشاهدة الجديدة. ويتعين على الباحث عدم المبالغة في التنبؤ باستخدام قيم تبعد كثيراً عن الحد الأدنى والأقصى لقيم المتغير المستقل في العينة التي استخدمت في بناء النموذج.

فترة التنبؤ:

لإيجاد فترة التنبؤ للقيمة المتنبأ بها يجب ملاحظة أنه يوجد مصدران للخطأ هما: الخطأ الفردي كما هو مقاس بـ (σ^2) وخطأ تقدير القيمة المتوقعة ($E(Y/X_0)$) باستخدام ($\hat{y}_0$)، ويمكن تمثيل هذه الأخطاء بالمعادلة التالية:

$$y-\hat{y}_o = y-E\left(y/x_0\right)+E\left(y/x_0\right)-\hat{y}_o$$

وبما أن المشاهدة الجديدة مستقلة عن العينة الأصلية، فأنه يتم إيجاد تباين ($y-\hat{y}_o$) كما يلي:

$$\text{var}\left(\hat{y}_h\right)=\text{var}\left(y-\hat{y}_o\right)$$

حيث إن $\hat{y}_h$ قيمة التنبؤ الفردي لـ y و $\hat{y}_0$ القيمة المتنبأ بها عند x_0.

$$\text{var}(\hat{y}_h)=\text{var}(y)+\text{var}(\hat{y}_o)=\sigma^2+\sigma^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

وبما أن قيمة (σ^2) مجهولة فإنه يستعاض عنها بالمقدر غير المتحيز (S^2) وبالتالي يأخذ مقدر تباين التنبؤ الفردي الصيغة التالية:

$$S^2(\hat{y}_h)=S^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

وباتباع نفس الخطوات التي اتبعت لتأسيس فترة الثقة للقيمة المتوسطة، نجد أن فترة التنبؤ للتنبؤ الفردي هي:

$$\hat{\beta}_0+\hat{\beta}_1x_0\pm t_{n-2,\alpha/2}\,s.e(\hat{y}_h) \qquad (2\text{-}67)$$

حيث إن:

$$s.e(\hat{y}_h) = \sqrt{S^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)}$$

الانحراف المعياري للتنبؤ الفردي.

مثال:

استخدم بيانات نموذج انحدار وزن الطفل لتقدير والتنبؤ بوزن الطفل عندما يكون عمره (٦,٥) سنة ومن ثم احسب فترتي الثقة والتنبؤ باستخدام معامل ثقة (٩٥%).

الحل:

بما أن قيمة المتغير المستقل المراد عندها التقدير والتنبؤ هي ($x_0 = 6.50$)، فإن:

$$\hat{y}_o = \hat{\beta}_0 + \hat{\beta}_1 x_o = 4.210 + 2.53204 \times 6.5 = 20.67\ kg$$

أي أن وزن الطفل سيبلغ ٢٠,٦٧ كيلوجرام إذا بلغ عمره (٦,٥) عام ويتساوى ذلك في حالة تقدير القيمة المتوسطة والتنبؤ الفردي.

فترة ثقة (٩٥%) للقيمة المتوسطة:

بالتعويض في المعادلة (2.66) نتحصل على فترة الثقة لوزن الطفل البالغ (20.67) كجم كما يلي:

$$\hat{y}_0 \pm t_{n-2,\alpha/2} s.e(\hat{y}_0)$$

$$20.67 \pm 2.0106 \times 2.32446674\left(\frac{1}{50}+\frac{(6.5-2.3472)^2}{216.218408}\right) = 20.67 \pm 2.0106 \times 0.482$$

أو

$$(19.70 < \hat{y}_o < 21.64)$$

أي أن فترة ثقة ٩٥% لمتوسط وزن الطفل عندما يكون عمره (٦,٥) أعوام تمتد من ١٩,٧٠ إلى ٢١,٦٤ كيلوجرام.

أما فترة ثقة التنبؤ الفردي فإنه يتم حسابها كما يلي:

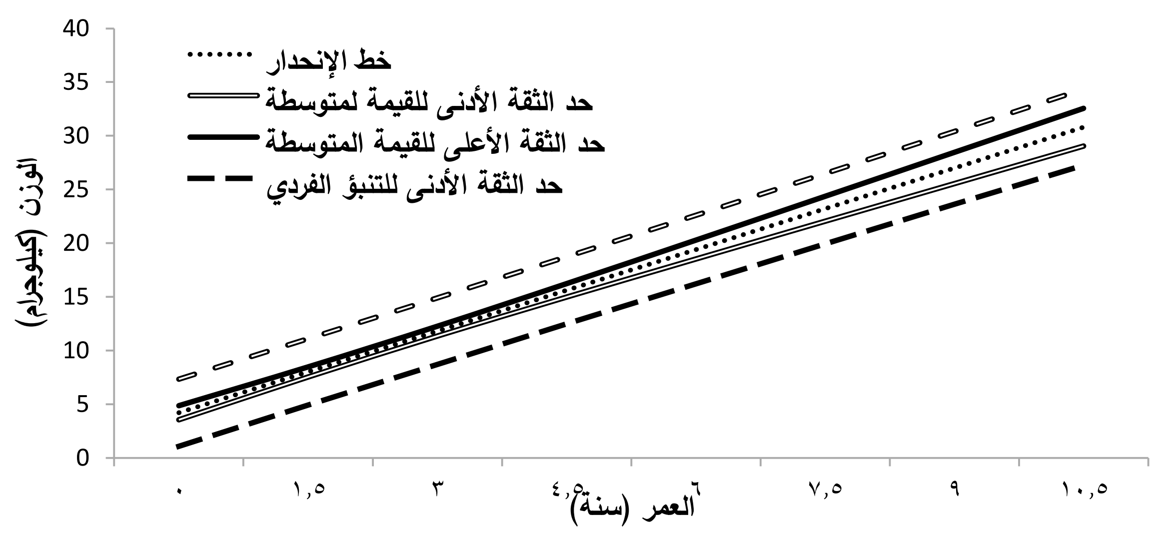
$$\hat{y}_h \pm t_{n-2,\alpha/2}\sqrt{S^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)}$$

$$20.67 \pm 2.0106\sqrt{2.32446674(1+\frac{1}{50}+\frac{(6.5-2.3472)^2}{216.2184808})}$$

أو

$$(17.46 < \hat{y}_h < 23.89)$$

أي أن فترة تنبؤ ٩٥% الخاصة بوزن الطفل عندما يكون عمره (٦,٥) عام تمتد من ١٧,٤٦ إلى ٢٣,٨٩ كيلوجرام. ويلاحظ أن فترة التنبؤ التي حصلنا عليها أكثر اتساعًا من فترة الثقة للقيمة المتوسطة. والشكل رقم (٢-١٢) يوضح فترتي الثقة للقيمة المتوسطة والتنبؤ الفردي لنموذج انحدار وزن الطفل على عمره عند قيم مختارة للمتغير المستقل عند فترة ثقة (٩٥%).



شكل رقم (٢-١٢): فترات الثقة والتنبؤ عند فترة ثقة (٩٥%) لوزن الطفل المناظرة لقيم مختارة لعمر الطفل

٢-١٢ تحويل بعض نماذج الانحدار غير الخطية إلى نماذج خطية:

قد تربط أحيانًا بين المتغير التابع والمتغير المستقل علاقة غير خطية المعالم أو غير خطية المتغيرات أو غير خطية المعالم والمتغيرات معًا. وتقترح بعض النظريات العلمية وجود علاقات غير خطية. فمثلًا يستخدم الاقتصاديون دالة إنتاج كوب-دوجلاس (Cobb-Douglas Production Function) لقياس العلاقة بين الإنتاج وكل من رأس المال والعمالة التي تأخذ الصيغة غير الخطية التالية:

$$Y_i = \beta_0 L_i^{\beta_1} K_i^{\beta_2} e^{\varepsilon_i} \qquad (2-68)$$

حيث إن Y الإنتاج، L العمالة، K رأس المال e أساس اللوغاريتم الطبيعي و ε حد الخطأ العشوائي.

إلا أنه لا يوجد دائمًا نماذج نظريات توضح شكل العلاقة بين المتغير التابع والمتغير/المتغيرات المستقلة؛ وفي هذه الحالة يستعين الباحث برسم شكل الانتشار لتحديد شكل الدالة التي توفق البيانات. تستخدم طريقة المربعات الصغرى لتقدير معالم بعض النماذج غير الخطية وذلك بعد إجراء تحويلات مناسبة للمتغيرات بحيث تأخذ العلاقة المعطاة الصورة الخطية. وكما أشرنا سابقًا إلى أنه يستخدم شكل الانتشار قبل البدء في بناء نموذج الانحدار للكشف عن شكل الدالة المناسب. فإذا ما تلاحظ أن العلاقة بين المتغيرين التابع والمستقل علاقة غير خطية فإنه يجب اختيار تحويلة أو تحويلات مناسبة للمتغير التابع أو المتغير المستقل أو للمتغيرين معًا بحيث تأخذ العلاقة الصورة الخطية.

لقد طوّر كل من بوكس وكوكس (Box and Cox (1964) pp. 211-252) مجموعة من التحويلات تعرف بتحويلات القوة للتغلب على بعض المشكلات مثل عدم تبعية حد الخطأ العشوائي للتوزيع الطبيعي، عدم ثبات تباين حد الخطأ العشوائي وعدم خطية دالة الانحدار. وتأخذ هذه التحويلات الصيغة التالية:

$$Y' = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \ln Y & \text{for } \lambda = 0 \end{cases} \qquad (2.69)$$

حيث إن:

Y' = قيمة المتغير Y بعد إجراء التحويلة له.

λ = معلمة تأخذ قيمًا سالبة أو موجبة.

Ln= اللوغاريتم الطبيعي للأساس e (e=2.718)

ويلاحظ من المعادلة (2.69) أن تحويلة بوكس-كوس هي في الأساس تحويلة قوة تأخذ الصيغة ($Y' = Y^{\lambda}$) لكل قيم λ الموجبة والسالبة ما عدا عندما تساوي الصفر عندها تكون التحويلة تحويلة لوغاريتم طبيعي وذلك لأن:

$$\lim_{\lambda \to 0} \left(\frac{Y^{\lambda} - 1}{\lambda} \right) = \ln Y$$

ويُعَدّ رسم شكل الانتشار بين المتغير التابع والمتغير المستقل هو الخطوة الأولى لتحديد قيمة مناسبة لـ λ. ويقترح نيتر وآخرون (Neter et al, 1990 p. 149) أن يقوم الباحث مستعيناً برسم شكل الانتشار باختيار عدة قيم لـ λ لإجراء تحويلات مناسبة ومن ثم يتم بناء نماذج انحدار خطي بعدد التحويلات ومن بعد ذلك يتم اختيار النموذج الذي لديه أقل مجموع مربعات بواقي.

٢-١٢-١ بعض التحويلات المستخدمة:

فيما يلي نستعرض أكثر التحويلات المستخدمة واستعمالاتها:

التحويلة اللوغاريتمية ($\ln Y$):

وتستخدم التحويلة اللوغاريتمية في الحالات التالية:

- تثبيت تباين المتغير التابع إذا اتضح أنه يتزايد بزيادة قيم المتغير المستقل.
- للحصول على اعتدالية المتغير التابع إذا ما بدا لنا من شكل الانتشار أن توزيعه ملتو نحو اليمين (موجب الالتواء).
- تحويل دالة غير خطية إلى دالة خطية.

تحويلة التربيع (Y^2):

وتستخدم تحويلة التربيع في الحالات التالية:

- تثبيت تباين حد المتغير التابع إذا اتضح أنه يتناقص مع الوسط الحسابي للمتغير التابع.
- للحصول على اعتدالية المتغير التابع إذا ما بدا لنا من شكل الانتشار أن توزيعه ملتو نحو اليسار (سالب الالتواء).
- تحويل دالة غير خطية إلى دالة خطية.

تحويلة الجذر التربيعي للمتغير التابع ($\sqrt{Y}$):

- وتستخدم هذه التحويلة لتثبيت تباين المتغير التابع إذا كان يتناسب مع الوسط الحسابي لـ Y. وتعتبر هذه التحويلة مناسبة إذا كان المتغير التابع له توزيع بواسون (Poisson Distribution).

٢-١٢-٢ بعض النماذج غير الخطية وكيفية تحويلها إلى خطية:

أ- نماذج اللوغاريتم المزدوج (Double –log Models):

يأخذ نموذج اللوغاريتم المزدوج الذي يعرف أيضًا بالنموذج الأسي الصيغة التالية:

$$y_i = \beta_0 x_i^{\beta_1} e^{\varepsilon_i} \qquad (2\text{-}70)$$

ويلاحظ أن منحنى الدالة في هذا النموذج يمر بنقطة الأصل (نقطة تساوي قيم كل من X وY بالصفر) وأن ميل المنحنى يتزايد أو يتناقص حسب قيمة المعلمة β_1، فإذا كانت β_1 أكبر من الواحد الصحيح فإن الميل يزيد بزيادة X، أما إذا كانت قيمة β_1 أقل من الواحد الصحيح فإن الميل يتناقص عندما تزيد قيم X (انظر الشكلين (٢-١٣) و(٢-١٤)).

ولتقدير معالم نموذج الانحدار الأسي باستخدام طريقة المربعات الصغرى يتم تحويل المعادلة (2.70) إلى شكل خطي وذلك بأخذ لوغاريتم طرفي المعادلة كما يلي:

$$\ln y_i = \ln \beta_0 + \beta_1 \ln x_i + \varepsilon_i \qquad (2\text{-}71)$$

ومن المزايا التي جعلت نموذج اللوغاريتم المزدوج أكثر شيوعًا وتطبيقًا هو أن المعامل β_1 يقيس مرونة المتغير التابع (Y) بالنسبة للمتغير المستقل (X)، أي زيادة نسبة مئوية واحدة في المتغير المستقل X تصاحبها تغير نسبي ثابت في القيمة المتوسطة للمتغير التابع Y.

ب- النماذج شبه اللوغاريتمية (Semi-log Models):

تسمى هذه النماذج بشبه اللوغاريتمية لأن أحد المتغيرين يتم تحويله إلى لوغاريتم لكي يأخذ النموذج الشكل الخطي.

- النموذج الأول (لوغاريتم المتغير التابع): يأخذ هذا النموذج الذي يعرف أيضًا بالنموذج الأسي، الصيغة التالية:

$$y_i = e^{\beta_0 + \beta_1 x_i + \varepsilon_i} \qquad (2\text{-}72)$$

ولهذا النموذج، كما يلاحظ، ثابت موجب (المقطع الصادي) يساوي e^{β_0} (انظر الشكلين (٢-١٥) و(٢-١٦). ومن خصائص هذا النموذج أن الميل (المعامل β_1) يقيس التغير النسبي في المتغير التابع (Y) الذي يصاحب التغير المطلق في المتغير المستقل (X)، أي أن:

$$\beta_1 = \frac{\text{التغير النسبي في Y}}{\text{التغير المطلق في X}}$$

ويستخدم هذا النموذج (2.72) لتقدير معدلات النمو إذا كان المتغير المستقل يمثل اتجاهًا زمنيًا، مثال ذلك نمو السكان، الصادرات.. إلخ بمعدل ثابت عبر الزمن. ولتحويل هذا النموذج إلى نموذج خطي نأخذ لوغاريتم طرفي المعادلة (2.72) كما يلي:

$$\ln y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (2\text{-}73)$$

وبالتالي يمكننا استخدام طريقة المربعات الصغرى لتقدير معالم النموذج الأسي.

- النموذج الثاني (لوغاريتم المتغير المستقل): يأخذ هذا النموذج الصيغة التالية:

$$e^{y_i} = \beta_0 x_i^{\beta_1} e^{\varepsilon_i} \qquad (2\text{-}74)$$

ويستخدم هذا النموذج إذا كان التغير في المتغير المستقل (X) بنسبة ثابتة يؤدي إلى تغير في المتغير التابع بمقدار ثابت انظر الشكلين (٢-١٧) و(٢-١٨). وفي هذا النموذج يقيس الميل التغير المطلق في المتغير التابع الذي يصاحب التغير النسبي في المتغير المستقل (X)، أي أن:

$$\beta_1 = \frac{\text{التغير المطلق في } Y}{\text{التغير النسبي في } X}$$

ولتقدير معالم النموذج (2.74) باستخدام طريقة المربعات الصغرى يتم تحويله إلى نموذج خطي وذلك بأخذ لوغاريتم طرفي المعادلة كما يلي:

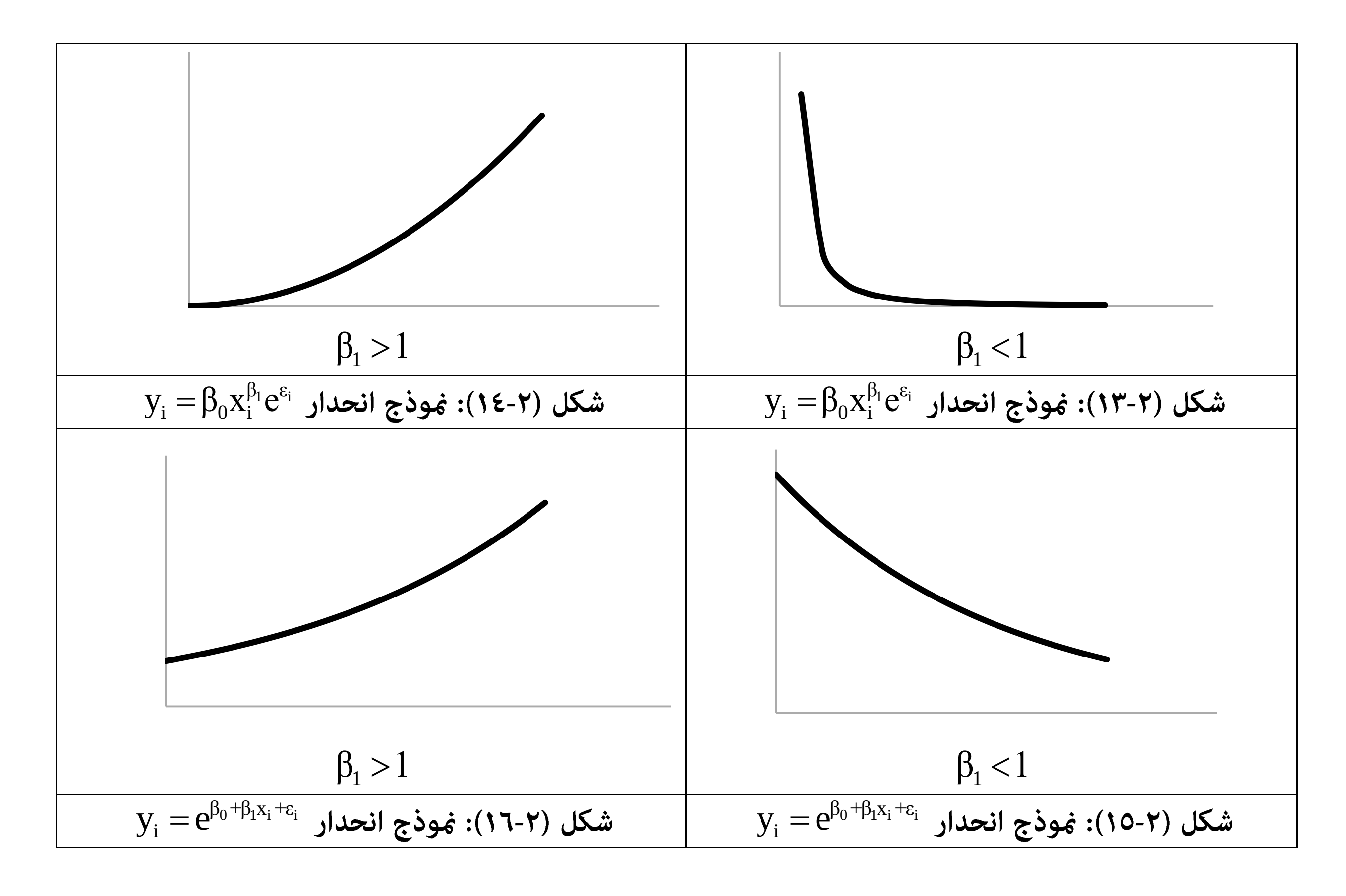
$$y_i = \ln\beta_0 + \beta_1 \ln x_i + \varepsilon_i \qquad (2\text{-}75)$$

ج- نموذج المقلوب (Hyperbolic or Reciprocal Model):

يأخذ نموذج المقلوب صيغًا مختلفة إلا أن أهمها وأكثرها بساطة يأخذ الصيغة التالية:

$$y_i = \beta_0 + \beta_1(\tfrac{1}{x_i}) + \varepsilon_i \qquad (2\text{-}76)$$

ويوضح الشكلان رقم (٢-١٩) و(٢-٢٠) منحنى النموذج في حالة β_1 موجبة وفي حالة β_1 سالبة. وأهم ما يميز المنحنى المقلوب هو كلما كانت قيمة X كبيرة وتؤول إلى ما لا نهاية فإن القيمة المتوقعة لـ Y تقترب إلى قيمة وإذا كانت قيمة β_1 سالبة فإن القيمة المتوقعة لـ Y تكون دائمًا أقل من β_0.

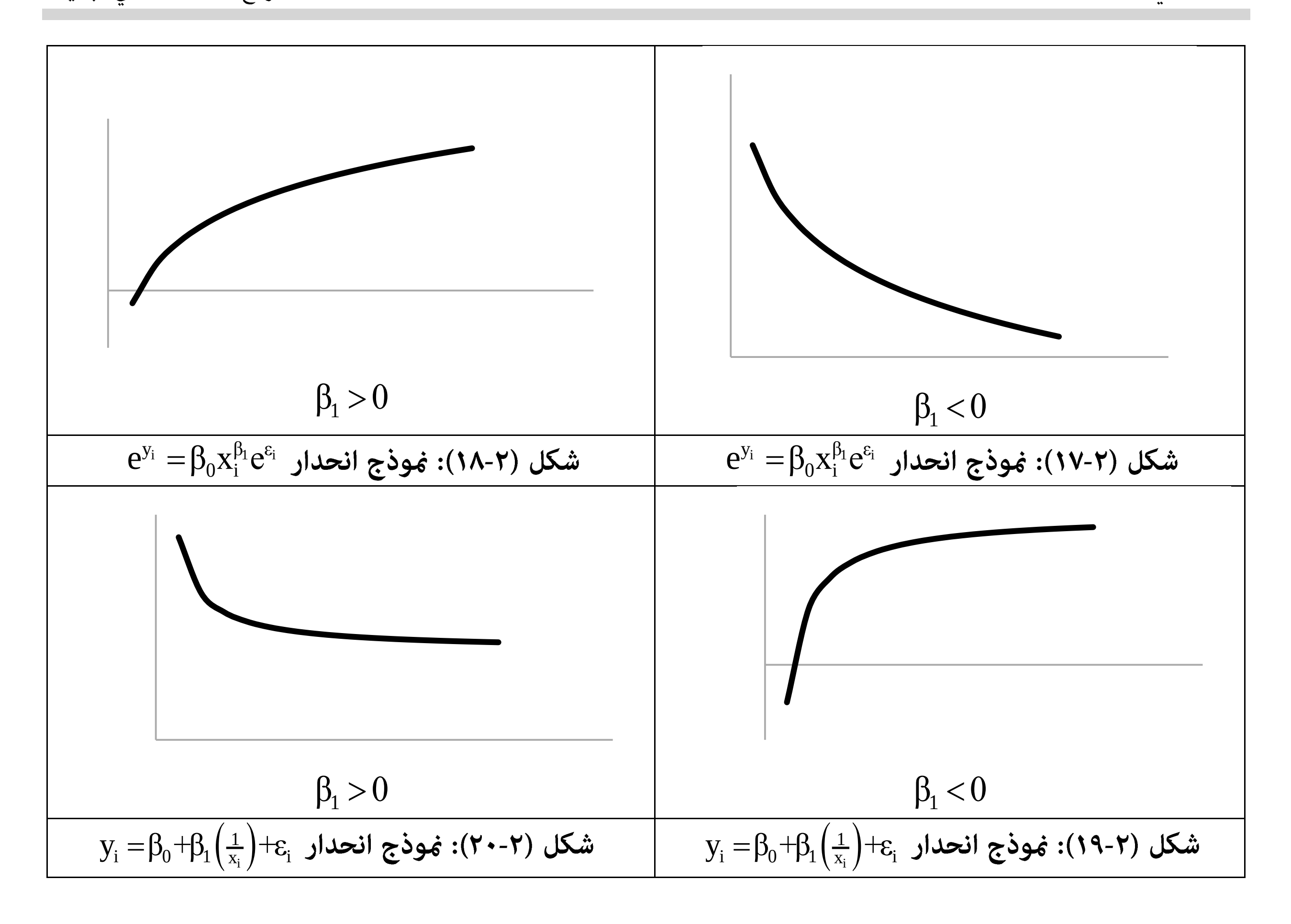


شكل (٢-١٣): نموذج انحدار $y_i = \beta_0 x_i^{\beta_1} e^{\varepsilon_i}$

شكل (٢-١٤): نموذج انحدار $y_i = \beta_0 x_i^{\beta_1} e^{\varepsilon_i}$

شكل (٢-١٥): نموذج انحدار $y_i = e^{\beta_0 + \beta_1 x_i + \varepsilon_i}$

شكل (٢-١٦): نموذج انحدار $y_i = e^{\beta_0 + \beta_1 x_i + \varepsilon_i}$



شكل (٢-١٧): نموذج انحدار $e^{y_i} = \beta_0 x_i^{\beta_1} e^{\varepsilon_i}$

شكل (٢-١٨): نموذج انحدار $e^{y_i} = \beta_0 x_i^{\beta_1} e^{\varepsilon_i}$

شكل (٢-١٩): نموذج انحدار $y_i = \beta_0 + \beta_1 \left(\frac{1}{x_i}\right) + \varepsilon_i$

شكل (٢-٢٠): نموذج انحدار $y_i = \beta_0 + \beta_1 \left(\frac{1}{x_i}\right) + \varepsilon_i$

مثال: نموذج نمو (Growth Model):

يبين الجدول رقم (٢-٦) إجمالي عرض النقود (النقود وشبه النقود) بالجمهورية اليمنية من العام ١٩٨٢م إلى ١٩٩٥م. ويظهر من الجدول الزيادة الكبيرة في عرض النقود خلال هذه الفترة، حيث ارتفع عرض النقود من حوالي ١٧ مليار ريال يمني في عام ١٩٨٢م إلى حوالي ٢٦٥,٧ مليار في عام ١٩٩٥م، أي بزيادة قدرها ١٤٧١%. في هذا المثال نود بناء نموذج انحدار خطي بسيط لقياس معدل النمو السنوي لعرض النقود. وبرسم شكل انتشار عرض النقود (متغير تابع) مع الزمن (متغير مستقل) نجد أن اتجاه عرض النقود يأخذ شكل منحنى أسي (شكل رقم (٢-٢١))، ولذلك من المناسب في هذه الحالة توفيق نموذج أسي يأخذ الصيغة التالية:

$$y_i = e^{\beta_0 + \beta_1 x_i + \varepsilon_i}$$

حيث إن Y يمثل عرض النقود وX الزمن.

ولكي نستخدم طريقة المربعات الصغرى لتقدير معالم النموذج لا بـد مـن تحويـل العلاقـة إلى خطيـة وذلـك بأخـذ لوغاريتم طرفي المعادلة على النحو التالي:

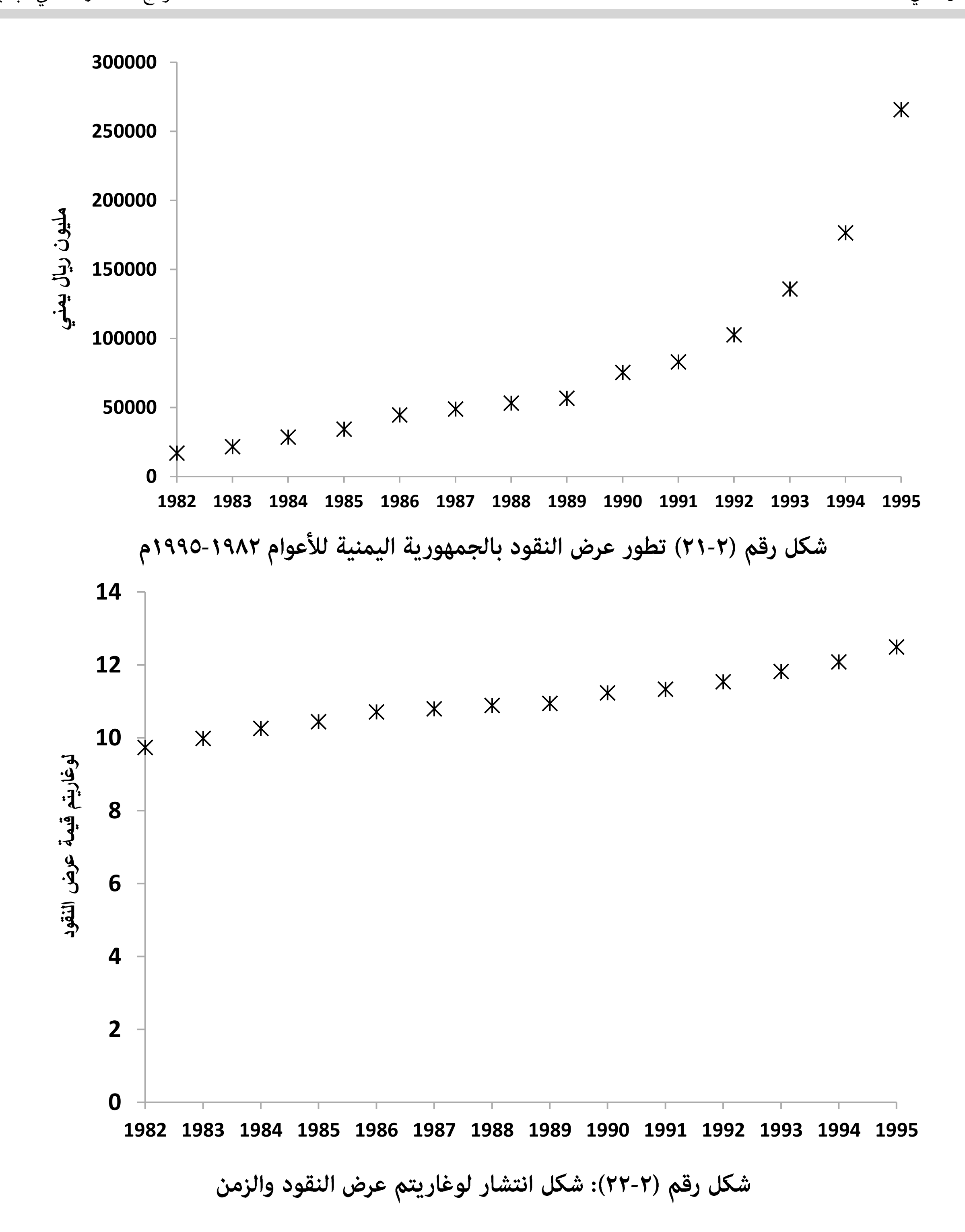
$$\ln y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

وبإعادة رسم شكل انتشار عرض النقود بعد تحويله بأخذ اللوغاريتم مع الزمن نجد أن العلاقة قد أصـبحت خطيـة (شكل رقم (٢-٢٢)).

جدول رقم (٢-٦): عرض النقود (مليون ريال يمني) بالجمهورية اليمنية للأعوام (١٩٨٢-١٩٩٥)

بنهاية العام الميلادي	عرض النقود (النقد + شبه النقد)	لوغاريتم عرض النقود
١٩٨٢	١٦٩١٦,١	٩,٧٣٦
١٩٨٣	٢١٥٩٧,٥	٩,٩٨٠
١٩٨٤	٢٨٥٦٨,٧	١٠,٢٦٠
١٩٨٥	٣٤٢٦٧,٥	١٠,٤٤٢
١٩٨٦	٤٤٦٣١,٢	١٠,٧٠٦
١٩٨٧	٤٨٨٤٧,٢	١٠,٧٩٦
١٩٨٨	٥٣٢٥٣,٩	١٠,٨٨٣
١٩٨٩	٥٦٦٨٠,٥	١٠,٩٤٥
١٩٩٠	٧٥٤٤٩,٤	١١,٢٣١
١٩٩١	٨٣١٤٣,٦	١١,٣٢٨
١٩٩٢	١٠٢٥٩٠,١	١١,٥٣٩
١٩٩٣	١٣٥٨٦٦,٣	١١,٨١٩
١٩٩٤	١٧٦٥٣٦,٨	١٢,٠٨١
١٩٩٥	٢٦٥٧٣٩,٦	١٢,٤٩٠

المصدر: عبد المعطي محمد عساف (١٩٩٧) ص ص ٤٨٣-٥٤٠



شكل رقم (٢-٢١) تطور عرض النقود بالجمهورية اليمنية للأعوام ١٩٨٢-١٩٩٥م

شكل رقم (٢-٢٢): شكل انتشار لوغاريتم عرض النقود والزمن

نتائج النموذج:

توضح النتائج المستعرضة بالإطار رقم (٢-١) أن هناك علاقة إيجابيه ذات دلالة إحصائية بين لوغاريتم عرض النقود والزمن (P-value <0.01) وأن نموذج الانحدار الذي يوضح هذه العلاقة هو:

$$\ln \hat{y}_i = 9.614 + 0.187 \times \text{TIME}$$

كما يمكن كتابة هذا النموذج بالصيغة الأسية كما يلي:

$$\hat{y}_i = e^{9.614 + 0.187 \times \text{Time}}$$

ويفسر هذا النموذج ٩٧,٧% من تباين لوغاريتم عرض النقود. ويشير ميل النموذج إلى أن معدل النمو السنوي لعرض النقود خلال الفترة من ١٩٨٢ إلى ١٩٩٥م قد بلغ ١٨,٧%. وباستخدام هذا النموذج يمكننا التنبؤ بقيم عرض النقود للعام ١٩٩٦ وذلك بالتعويض في المعادلة أعلاه ومن ثم إيجاد عكس اللوغاريتم (Anti-log) للقيم المتنبأ بها. حيث يتم حساب القيمة المتنبأ بها لعرض النقود للعام ١٩٩٦م كما يلي:

$$\hat{y}_{1996} = 9.614 + 0.187 \times 15 = 12.4198$$

ويتم حساب فترة تنبؤ ٩٥% للقيمة المتنبأ بها للعام ١٩٩٦م كما يلي:

$$12.4198 \pm 2.179x\sqrt{0.0159(1 + \frac{1}{14} + \frac{(15-7.5)^2}{227.5}} = 12.4198 \pm 0.3155309$$

أو

$$12.1043 < \hat{y}_{1996} < 12.7353$$

أي أننا واثقون بدرجة ٩٥% أن تكون القيمة المتنبأ بها ما بين ١٢,١٠٤٣ و١٢,٧٣٥٣. وبإيجاد معكوس اللوغاريتم للقيمة المتنبأ بها وفترة التنبؤ يمكننا القول بأن عرض النقود بالجمهورية اليمنية سيبلغ ٢٤٧٦٥٧ ألف ريال يمني بنهاية

العام ١٩٩٦م، وأن فترة الثقة (٩٥%) الخاصة بهذا التنبؤ الفردي تمتد من ١٨٠٦٤٧ حدًّا أدني و٣٣٩٥٢٤ ألف ريال يمني حدًّا أعلى.

إطار رقم (٢-١): جزء من مخرجات نتائج نموذج انحدار لوغاريتم عرض النقود على الزمن باستخدام برنامج Excel

Regression Statistics	
R Square	0.976602925
Standard Error	0.126053788
Observations	14

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7.958835	7.958835	500.8846	0.0000
Residual	12	0.190675	0.01589		
Total	13	8.14951			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	9.614204886	0.07116	135.1074	0.0000	9.459161187	9.769249
الزمن	0.187039809	0.008357	22.38045	0.0000	0.16883085	0.205249

تمارين

١) يوضح الجدول التالي بيانات افتراضية عن الأداء الوظيفي (Y) والمستوى التعليمي (X) لعدد (٢٥) من منسـوبي إحدى الشركات.

مسلسل	الأداء الوظيفي (من ١٠٠درجة)	عدد سنوات التعليم
1	92	19
2	93	19
3	94	18
4	87	15
5	88	15
6	75	15
7	55	9
8	63	12
9	62	12
10	48	8
11	97	19
12	63	14
13	58	9
14	61	11
15	83	18
16	75	17
17	71	16
18	70	17
19	65	12
20	63	11
21	54	8
22	87	18
23	55	13
24	66	12
25	77	17

المطلوب:

- ارسم شكل انتشار الأداء الوظيفي وعدد سنوات التعليم.
- هل العلاقة خطية؟ إذا كانت خطية قدِّر معادلة انحدار الأداء الوظيفي على عدد سنوات التعليم.
- اختبر دلالة مقدرات معالم النموذج باستخدام مستوى دلالة إحصائية (١%).
- احسب معامل الارتباط الخطي البسيط ومعامل التحديد.
- احسب جدول تحليل التباين.
- كون فترة ثقة ١% لمعاملي الانحدار.
- فسر النتائج التي تحصل عليها من الأسئلة أعلاه.

٢) افترض باحث وجود علاقة بين مستوى ضغط الدم وعمر الفرد، ولاختبار هذه الفرضية جمع الباحث بيانات حول هذين المتغيرين من (٤٣) مريضًا ووجد أن معامل الارتباط بين هذين المتغيرين قد بلغ ($r=0.5236$).

المطلوب:

- هل هذه العلاقة دالة إحصائيَا؟
- بناءَ على النتيجة التي ستحصل عليها هل يمكننا القول بأن هناك علاقة خطية بين مستوى ضغط الدم وعمر الفرد؟

٣) جدول تحليل التباين التالي هو جزء من نتائج نموذج انحدار مبيعات سلعة ما على متغير مصروفات الدعاية والإعلان عليها.

مصدر التغير/التباين	درجات الحرية	مجموع المربعات	متوسط المربعات	قيمة إحصائية F_0
الانحدار (التباين المفسر)	-	١٣٧١٨٨٢,١	-	-
البواقي (التباين غير المفسر)	١٨	-	-	
المجموع (التباين/التغير الكلي)	-	١٦٥٠٠٠٠		

المطلوب:

- احسب القيم المشار إليها بعلامة "-" بالجدول.
- هل العلاقة بين المبيعات ومصروفات الدعاية والإعلان دالة إحصائيَا؟

٤) يُفترض أحياناً أن خط الانحدار يمر بنقطة الأصل (Regression through the origin)، أي أن النموذج لا يضم في هذه الحالة المعامل الثابت (Intercept)، حيث يأخذ النموذج الصيغة التالية:

$$Y_i = \beta X_i + \varepsilon_i \qquad where\ \varepsilon_i \sim NID\left(0, \sigma^2\right)$$

- أوجد مقدر المربعات الصغرى للمعلمة β ($\hat{\beta}$) وبرهن على أن هذا المقدر غير متحيز.
- أوجد تباين المقدر ($\hat{\beta}$).
- اشتق إحصائية اختبار الفرض التالي:

فرض العدم: $H_0 : \beta = 0$ مقابل الفرض البديل: $H_1 : \beta \neq 0$

٥) باستخدام طريقة الإمكان الأعظم (MLE)، اشتق مقدر التباين لنموذج الانحدار الخطي البسيط، هل مقدر التباين مقدر غير متحيز؟ .

الفصل الثالث

نموذج الانحدار الخطي المتعدد

٣-١ مقدمــة:

اقتصرت دراستنا لتحليل الانحدار الخطي البسيط على تحليل أثر متغير مستقل واحد على المتغير التـابع. إلا أنـه في الواقع نادراً ما نجد متغيراً واحداً يفسر جزءاً كبيراً من التغير أو التباين في المتغير التابع. فمثلاً من غير المتوقع أن يكـون الصرف على دعاية وترويج سلعة ما هو المتغير الوحيد الذي يـؤثر في مبيعـات السـلع، بـالطبع توجـد متغـيرات أخـرى مؤثرة كالسعر، هامش الربح، ذوق المستهلك، أسعار السلع البديلة، ... إلخ. ولـذلك نجـد أن نمـوذج الانحـدار الخطـي المتعدد الذي يقيس أثر أكثر من متغير واحد على المتغير التابع هو الأوسع استخداماً. ويعتـبر نمـوذج الانحـدار الخطـي المتعدد تعميماً للمفاهيم والأساليب والصيغ المستخدمة في نموذج الانحـدار الخطـي البسـيط الـذي سـبق مناقشـته في الفصل الثاني. ويتمثل الفرق الوحيد بين نموذجي الانحدار البسيط والانحدار المتعـدد في أن الأول يضـم متغـيراَ مسـتقلاً واحدًا في حين يضم الثاني متغيرين مستقلين أو أكثر من المتغيرات المستقلة. والجدول رقم (٣-١) يوضـح بيانـات نمـوذج الانحدار المتعدد المتكونة من (n) من المشاهدات للمتغير التابع مع (p) من المتغيرات المستقلة.

جدول رقم (٣-١): بيانات نموذج الانحدار الخطي المتعدد

رقم المشاهدة	المتغير التابع	المتغيرات المستقلة			
	y_i	x_{1i}	x_{2i}	. . .	x_{pi}
1	y_1	x_{11}	x_{21}	. . .	x_{p1}
2	y_2	x_{12}	x_{22}	. . .	x_{p2}
3	y_3	x_{13}	x_{23}	. . .	x_{p3}
.	.	.	.	.	.
n	y_n	x_{1n}	x_{2n}	. . .	x_{pn}

يعالج هذا الفصل موضـوعات تحليـل الانحـدار الخطـي المتعـدد مـن حيـث بنـاء النمـوذج والتقـدير والاختبـارات الإحصائية والتنبؤ والحزم الإحصائية الأكثر اسـتخداماً لإجـراء تحليـل الانحـدار. وسنسـتخدم في هـذا الفصـل والفصـول اللاحقة طرق الجبر الخطي (Linear Algebra) لحساب مقدرات المربعـات الصـغرى والإحصـاءات الأخـرى المرتبطـة بهـا نظراً للسهولة التي تتميز بها في معالجة المتغيرات المتعددة.

٣-٢ نموذج الانحدار الخطي المتعدد:

تأخذ معادلة نموذج الانحدار الخطي المتعدد الذي يضم عدد (p) متغير مستقل الصيغة التالية:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i \qquad i = 1,2,\ldots,N \\ &= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji} + \varepsilon_i \end{aligned} \tag{3-1}$$

حيث إن:

y المتغير التابع.

$\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ معالم النموذج المجهولة المراد تقديرها وتسمى أيضاً بمعاملات الانحدار الجزئية (Partial regression coefficients).

$x_1, x_2, \ldots, x_p$ المتغيرات المستقلة.

N عدد المشاهدات (حجم المجتمع)

والمعادلة (3.1) هي اختصار لعدد N من المعادلات الآنية التالية:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_p x_{p1} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_p x_{p2} + \varepsilon_2 \\ y_3 &= \beta_0 + \beta_1 x_{13} + \beta_2 x_{23} + \ldots + \beta_p x_{p3} + \varepsilon_3 \\ &\vdots \\ y_N &= \beta_0 + \beta_1 x_{1N} + \beta_2 x_{2N} + \ldots + \beta_p x_{pN} + \varepsilon_N \end{aligned} \tag{3-2}$$

وباستخدام صيغ المصفوفات يمكن كتابة المعادلة (3.2) على النحو التالي:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & . & . & . & x_{p1} \\ 1 & x_{12} & x_{22} & . & . & . & x_{p2} \\ 1 & x_{13} & x_{23} & . & . & . & x_{p3} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & x_{1N} & x_{2N} & & & & x_{pN} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix} \tag{3-3}$$

كما يمكن كتابة المعادلات (3.1) و(3.2) و(3.3) باستخدام رموز المصفوفات كما يلي:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (3\text{-}4)$$

حيث إن $\mathbf{y}$ متجه عمودي من الدرجة $N\times1$ يحتوي على N مشاهدة للمتغير التابع، $\mathbf{X}$ مصفوفة البيانات (Data Matrix) من الدرجة $N\times p$ تحتوي على مشاهدات المتغيرات المستقلة $(x_1, x_2,\ldots,x_p)$ حيث يحتوي العمود الأول على قيم الواحد الصحيح لتمثيل المعامل الثابت (Intercept)، $\boldsymbol{\beta}$ متجه عمودي من الدرجة $p\times1$ يحتوي على معالم نموذج الانحدار المجهولة $\left(\beta_0,\beta_1,\beta_2,\ldots,\beta_p\right)$ المراد تقديرها، و $\boldsymbol{\varepsilon}$ متجه عمودي من الدرجة $N\times1$ يحتوي على قيم المتغير العشوائي (ε_i).

٣-٣ اشتراطات نموذج الانحدار الخطي المتعدد:

إن اشتراطات نموذج الانحدار الخطي المتعدد هي تعميم لاشتراطات نموذج الانحدار الخطي البسيط التي تمت مناقشتها في الفصل الثاني والاشتراط الإضافي الوحيد المطلوب لنموذج الانحدار المتعدد هو عدم وجود علاقة ارتباط خطي بين اثنين أو أكثر من المتغيرات المستقلة. وفيما يلي نستعرض هذه الاشتراطات:

١. عدم وجود أخطاء توصيف للنموذج (No specification errors) ويشمل:

- وجود علاقة خطية بين المتغير التابع (Y) والمتغيرات المستقلة $(X_1, X_2,\ldots,X_p)$.
- أن يتضمن نموذج الانحدار الخطي المتغيرات المستقلة التي تسهم في تفسير المتغير التابع، أي عند بناء النموذج يجب أن لا ندخل متغيرات مستقلة ليست لها تأثير على المتغير التابع وأن لا نجهل في نفس الوقت إدخال متغيرات مستقلة ذات تأثير عليه.

٢. أن تكون قياسات المتغير التابع والمتغيرات المستقلة دقيقة وصحيحة.

٣. أن يكون تباين أي متغير مستقل أكبر من الصفر، أي:

$$\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2 > 0 \qquad \text{for } j=1,2,\ldots,p$$

حيث إن x_{ij} المتغير المستقل رقم j و $\overline{x}_j$ الوسط الحسابي له وn حجم العينة وp عدد المتغيرات المستقلة.

٤. أن تكون المتغيرات المستقلة $(X_1, X_2,\ldots,X_p)$ غير عشوائية (Non-stochastic)، أي أنها تحتوي قيماً ثابتة في المعاينات المتكررة. ولكي تكون قيم المتغيرات المستقلة غير عشوائية يجب على الباحث التحكم فيها تجريبياً كما هو الحال في العلوم الطبيعية. غير أنه في مجالات كثيرة كالعلوم الإنسانية والاجتماعية وغيرها لا يتحقق هذا الاشتراط، إذ يتم جمع البيانات عن طريق المشاهدة باستخدام أسلوب المعاينة أو الحصر الشامل. ولذلك نجد أن المتغيرات المستقلة في هذه الحقول تأخذ قيماً عشوائية. وفي هذه الحالة يضاف اشتراط آخر وهو أن تكون

المتغيرات المستقلة مستقلة عن حد الخطأ العشوائي. ويرى كمينتا (Kmenta,1970,p301) أن عشوائية المتغيرات المستقلة لا تؤثر على الخصائص الحميدة لطريقة المربعات الصغرى- الطريقة التي سيتم استخدامها لتقدير معالم نموذج الانحدار الخطي البسيط أو المتعدد.

٥. الاشتراطات المتعلقة بحد الخطأ العشوائي(ε_i) :

- القيمة المتوقعة لحد الخطأ العشوائي تساوي صفراً ($\mathbf{E(\boldsymbol{\varepsilon}) = 0}$)، أي:

$$E\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ E(\varepsilon_3) \\ . \\ . \\ . \\ E(\varepsilon_N) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \qquad (3\text{-}5)$$

حيث $\boldsymbol{\varepsilon}$ متجه عمودي من الدرجة $N\times1$ يحتوي على حدود الخطأ و $\mathbf{0}$ متجه عمودي صفري من الدرجة $N\times1$. وينص هذا الاشتراط على أن القيمة المتوقعة لأي عنصر من عناصر المتجه $\boldsymbol{\varepsilon}$ تساوي صفراً.

- ثبات تباين حد الخطأ واستقلال قيم حدود الخطأ بعضها عن بعض، أي:

$$\begin{aligned} E(\varepsilon_i\varepsilon_j) &= 0 \qquad \text{for } i \neq j \\ &= \sigma^2 \qquad \text{for } i = j \qquad i,j = 1,2,\ldots,N \end{aligned} \qquad (3\text{-}6)$$

ويمكن كتابة المعادلة (3.6) بصيغ المصفوفات كما يلي:

$$\text{cov}(\boldsymbol{\varepsilon}) = E[(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))^T] = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I} \qquad (3\text{-}7)$$

حيث $\boldsymbol{\varepsilon}^T$ مبدلة المتجه العمودي $\boldsymbol{\varepsilon}$ و$\mathbf{I}$ مصفوفة وحدة من الدرجة $N\times N$. كما يمكن كتابة المعادلة (3.7) كما يلي:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = E\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix}\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & . & . & . & \varepsilon_N \end{pmatrix}$$

$$E(\boldsymbol{\varepsilon\varepsilon}^T) = E\begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \varepsilon_1\varepsilon_3 & . & . & . & \varepsilon_1\varepsilon_N \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \varepsilon_2\varepsilon_3 & . & . & . & \varepsilon_2\varepsilon_N \\ \varepsilon_3\varepsilon_1 & \varepsilon_3\varepsilon_2 & \varepsilon_3^2 & . & . & . & \varepsilon_3\varepsilon_N \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ \varepsilon_N\varepsilon_1 & \varepsilon_N\varepsilon_2 & \varepsilon_N\varepsilon_3 & . & . & . & \varepsilon_N^2 \end{pmatrix} \quad (3\text{-}8)$$

وبأخذ التوقع (E) لكل عنصر من عناصر المصفوفة السابقة نحصل على:

$$E(\boldsymbol{\varepsilon\varepsilon}^T) = E\begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & E(\varepsilon_1\varepsilon_3) & . & . & . & E(\varepsilon_1\varepsilon_N) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & E(\varepsilon_2\varepsilon_3) & . & . & . & E(\varepsilon_2\varepsilon_N) \\ E(\varepsilon_3\varepsilon_1) & E(\varepsilon_3\varepsilon_2) & E(\varepsilon_3^2) & . & . & . & E(\varepsilon_3\varepsilon_N) \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ E(\varepsilon_N\varepsilon_1) & E(\varepsilon_N\varepsilon_2) & E(\varepsilon_N\varepsilon_3) & . & . & . & E(\varepsilon_N^2) \end{pmatrix} \quad (3\text{-}9)$$

وباستخدام شرطي ثبات التباين ($E(\varepsilon_i^2) = \sigma^2$) وعدم وجود ارتباط ذاتي بين حدود الخطأ $0 = \text{cov}(\varepsilon_i\varepsilon_j)$ تأخذ المصفوفة (3.9) الصيغة التالية:

$$E(\boldsymbol{\varepsilon\varepsilon}^T) = \begin{pmatrix} \sigma^2 & 0 & 0 & . & . & . & 0 \\ 0 & \sigma^2 & 0 & . & . & . & 0 \\ 0 & 0 & \sigma^2 & . & . & . & 0 \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 0 & 0 & 0 & . & . & . & \sigma^2 \end{pmatrix} \quad (3\text{-}10)$$

وتعرف المصفوفة (3.10) بمصفوفة تباين- تغاير حد الخطأ العشوائي، حيث تمثل عناصرها القطرية التباين (σ^2) وجميع عناصرها خارج القطر مساوية للصفر وهي تمثل التغاير بين حدود الخطأ (Covariance).

- ولإجراء اختبارات الدلالة الإحصائية أو المعنوية يُشترط أن يتبع متجه حد الخطأ (ε) التوزيع الطبيعي المتعدد وسطه الحسابي متجه عمودي صفري $(\mathbf{0})$ وتباينه مصفوفة التباين والتغاير ($\sigma^2\mathbf{I}$)، أو اختصاراً

$$\boldsymbol{\varepsilon} \sim NID(\mathbf{0},\sigma^2\mathbf{I})$$

٦. أن يكون عدد المشاهدات أكبر من عدد المعالم المراد تقديرها، أي:

$$n > (p+1)$$

حيث إن n= عدد المشاهدات (حجم العينة)، وp=عدد المتغيرات المستقلة، و(p+1)=عدد معالم نموذج الانحدار بما في ذلك المعامل الثابت.

٧. أن تكون رتبة مصفوفة البيانات $\mathbf{X}$ كاملة (Full rank) أو مساوية لـ (p+1) (عدد الأعمدة في المصفوفة). وهذا يعني أن تكون أعمدة المصفوفة $\mathbf{X}$ مستقلة خطياً، أي أن لا يكون هناك عمود من المصفوفة يمكن كتابته كتركيب خطي (Linear combination) من الأعمدة الأخرى. ويعرف هذا الاشتراط بشرط عدم وجود ارتباط خطي تام بين المتغيرات المستقلة (No perfect multicollinearity).

٣-٤ تقدير معالم نموذج الانحدار الخطي المتعدد:

كما في تحليل الانحدار الخطي البسيط تستخدم طريقة المربعات الصغرى الاعتيادية لتقدير معالم نموذج الانحدار الخطي المتعدد. ويأخذ نموذج الانحدار المقدر من عينة (Sample regression function (SRF)) المقابل لمعادلة انحدار المجتمع (3.1) الصيغة التالية:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_p x_{pi} + e_i \qquad (3\text{-}11)$$

حيث إن:

$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ هي القيم المقدرة للمعالم $(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$

e_i الباقي وهو الفرق بين القيمة الفعلية للمتغير التابع والقيمة المقدرة لها للمشاهدة رقم (i)

n عدد المشاهدات (حجم العينة).

ويمكن كتابة المعادلة (3.11) في شكل المصفوفات كما يلي:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & . & . & . & x_{p1} \\ 1 & x_{12} & x_{22} & . & . & . & x_{p2} \\ 1 & x_{13} & x_{23} & . & . & . & x_{p3} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & x_{1n} & x_{2n} & . & . & . & x_{pn} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ . \\ \hat{\beta}_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_n \end{pmatrix}$$

ويلاحظ أن العمود الأول في مصفوفة البيانات يحتوي على القيمة واحد صحيح عند كل المشاهدات من (1) إلى (n) وذلك لتقدير المعامل الثابت، والعمود الثاني من المصفوفة يحتوي على قيم المتغير المستقل الأول (X_1)، وهكذا كل عمود يحتوي على قيم متغير مستقل محدد. وباستخدام رموز المصفوفات يمكن اختصار كتابة نموذج الانحدار كما يلي:

$$\mathbf{y} = \mathbf{x}\hat{\boldsymbol{\beta}} + \mathbf{e} \qquad (3\text{-}12)$$

حيث إن:

$\hat{\boldsymbol{\beta}}$ = متجه عمودي من الدرجة ($(p+1)\times 1$) يحتوي على قيم المعاملات المقدرة.

$\mathbf{X}$ = مصفوفة بيانات العينة من الدرجة ($n\times(p+1)$) حيث n عدد المشاهدات (حجم العينة).

$\mathbf{e}$ = متجه عمودي من درجة $n\times 1$ ويحتوي على البواقي.

وتستخدم طريقة المربعات الصغرى الاعتيادية كما سبق شرحها في الفصل الثاني لتقدير معالم نموذج الانحدار المتعدد التي نحصل عليها بتقليل أو تدنية مجموع مربعات البواقي ($\min\sum_{i=1}^{n} e^2_i$) إلى أدنى قيمة له. ويتم تقدير معالم نموذج الانحدار بحيث تكون الدالة

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \ldots - \hat{\beta}_p x_{pi}\right)^2 \qquad (3\text{-}13)$$

نهاية صغرى. ويمكن إعادة كتابة المعادلة (3.13) في شكل المصفوفات كما يلي:

$$\mathbf{e}^{T}\mathbf{e} = \left(e_1 \;\; e_2 \;\; e_3 \; . \; . \; . \; e_n\right)\begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + e_3^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2$$

وبما أن $\mathbf{y} = \mathbf{x}\hat{\boldsymbol{\beta}} + \mathbf{e}$ و $\mathbf{e} = \mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}$ فإن

$$\mathbf{e}^{T}\mathbf{e} = \left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)^{T}\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)$$

$$\mathbf{e}^{T}\mathbf{e} = \mathbf{y}^{T}\mathbf{y} - 2\hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y} + \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{x}\hat{\boldsymbol{\beta}} \qquad (3\text{-}14)$$

وبتفاضل المعادلة (3.14) بالنسبة إلى $\hat{\boldsymbol{\beta}}$ ومساواة ناتج التفاضل بالصفر نحصل على:

$$\frac{\partial \mathbf{e}^{T}\mathbf{e}}{\partial \widehat{\boldsymbol{\beta}}} = -2\mathbf{X}^{T}\mathbf{y} + 2\mathbf{X}^{T}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{0} \qquad (3\text{-}15)$$

إذن

$$\mathbf{X}^{T}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^{T}\mathbf{y} \qquad (3\text{-}16)$$

وبضرب قبلي لطرفي المعادلة (3.16) بـ $\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}$ نحصل على:

$$\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\mathbf{X}\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\mathbf{y}$$

إذن

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{y} \qquad (3\text{-}17)$$

وبذلك نحصل على مقدرات معالم النموذج شريطة أن تكون رتبة المصفوفة $\left(\mathbf{X}^{T}\mathbf{X}\right)$ كاملة، أي أن تكون غير مفردة (Non-singular) وذلك لإيجاد محددتها ومن ثم معكوسها $\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}$.

وللتحقق من أن مقدرات معالم النموذج $\widehat{\boldsymbol{\beta}}$ تحقق القيمة الدنيا لـ $\mathbf{e}^{T}\mathbf{e}$ يجب التأكد من أن مصفوفة هيسيان (Hessian matrix) مصفوفة موجبة محددة (positive definite matrix). وبتفاضل المعادلة (٣,١٥) للمتجه $\widehat{\beta}^{T}$ يتم الحصول على مصفوفة هيسيان التالية:

$$\frac{\partial^{2}\mathbf{e}^{T}\mathbf{e}}{\partial\widehat{\beta}\,\partial\widehat{\beta}^{T}} = 2\mathbf{X}^{T}\mathbf{X}$$

وبما أن الأعمدة في المصفوفة $\mathbf{X}$ مستقلة خطياً، فإن المصفوفة $2\mathbf{X}^{T}\mathbf{X}$ مصفوفة موجبة محددة ذلك لما يلي:

بافتراض أن $v \neq 0$، فإن

$$v^{T}\left(\mathbf{X}^{T}\mathbf{X}\right)v = \left(\mathbf{X}v\right)^{T}\mathbf{X}v = \|\mathbf{X}v\|_{2} > 0$$

٣-٥ تفسير معاملات الانحدار الجزئية:

كما في الانحدار الخطي البسيط يمثل المعامل الثابت (β_0) القيمة المتوسطة للمتغير التابع (Y) عندما تكون قيم المتغيرات المستقلة مساوية للصفر. ولكنَّ هناك ملاحظتين يجب أخذهما في الاعتبار عند تفسير المعامل الثابت سبق شرحهما عند دراستنا لنموذج الانحدار الخطي البسيط، هي:

- يجب أن يكون التنبؤ بقيم المتغير التابع (Y) بالتعويض في نطاق قيم مشاهدات المتغيرات المستقلة التي استخدمت في تقدير النموذج.

- يصعب تفسير المعامل الثابت إذا كانت قيمته سالبة وقيم المتغير التابع الفعلية موجبة.

بينما يمثل المعامل (β_1) التغير في القيمة المتوسطة للمتغير التابع الناتج عن تغير المتغير المستقل (X_1) بوحدة واحدة بافتراض ثبات قيم المتغيرات المستقلة الأخرى ($X_2,X_3,\ldots,X_p$). أو بمعنى آخر يقيس المعامل (β_1) الأثر المباشر أو الصافي لتغير (X_1) بوحدة واحدة على القيمة المتوقعة للمتغير التابع. وكذلك فإن المعامل (β_2) يمثل التغير في القيمة المتوقعة للمتغير التابع (Y) الناتج عن تغير X_2 بوحدة واحدة بافتراض ثبات قيم بقية المتغيرات المستقلة ($X_1,X_3,\ldots,X_p$)، وهكذا يستمر التفسير لبقية معاملات الانحدار الجزئية. كما أن لحجم وإشارات معاملات النموذج دلالات معينة، فالمعامل الموجب يشير إلى وجود علاقة طردية بين المتغير المستقل والمتغير التابع والمعامل السالب يشير إلى أن العلاقة بين المتغيرين علاقة عكسية. ويشير حجم المعامل إلى مقدار التغير الذي يحدث في المتغير التابع الناتج عن زيادة مقدارها وحدة واحدة في المتغير المستقل بافتراض ثبات قيم المتغيرات المستقلة الأخرى.

٣-٦ مثال:

نستخدم في هذا المثال بيانات أوزان (٥٠) طفلاً تم اختيارهم عشوائياً من سجلات مستشفى أبها بالمملكة العربية السعودية للنساء والتوليد، وهو نفس المثال الذي تم تناوله في بناء نموذج خطي بسيط في الفصل الثاني تضمن عمر الطفل متغيراً مستقلاً ووزنَ الطفل متغيراً تابعًا. وسنواصل في هذا المثال تحليل هذه البيانات ببناء نموذج انحدار خطي متعدد، وذلك بإدخال طول الطفل كمتغير مفسر آخر لوزن الطفل (جدول رقم ٣-٢). وعلى ذلك يمكننا بناء النموذج التالي:

$$\text{Weight}_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Height}_i + \varepsilon_i$$

والمراد تقديره بواسطة العلاقة المقدرة التالية:

$$\text{Weight}_i = \widehat{\beta}_0 + \widehat{\beta}_1 \text{Age}_i + \widehat{\beta}_2 \text{Height}_i + e_i$$

أو

$$y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i} + e_i \qquad i = 1,2,\ldots,50$$

حيث Y وزن الطفل (كيلو جرام)، X_1 عمر الطفل (سنة) وX_2 طول الطفل (سنتيمتر).

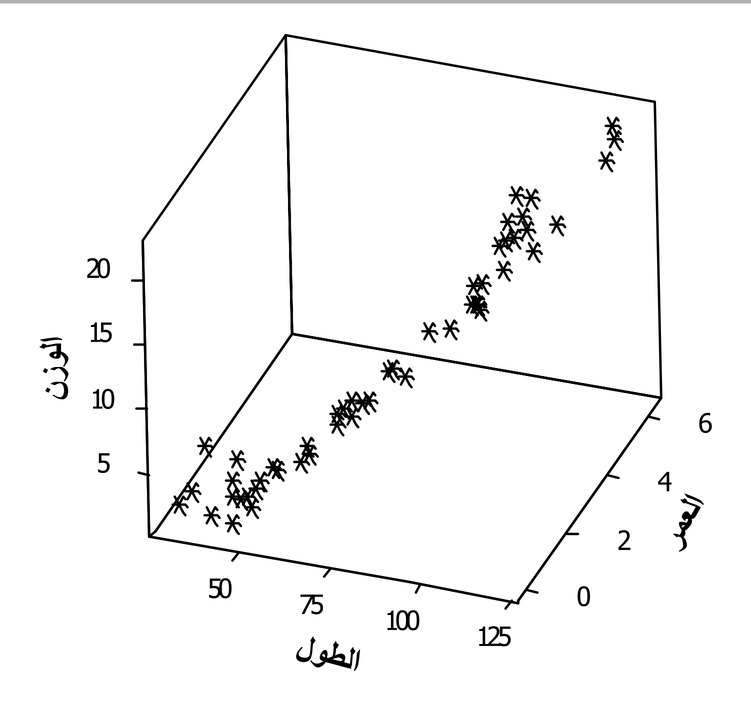
الحل:

نبدأ أولاً برسم شكل الانتشار للتأكد أولاً من خطية العلاقة بين وزن الطفل كمتغير تابع مع كل من متغيري الوزن والطول. ويوضح الشكل رقم (٣-١) رسم انتشار ثلاثي الأبعاد لمتغيرات الوزن والطول والعمر، ويوضح الشكل رقم (٣-٢) رسم انتشار الوزن مع كل من العمر والطول. ويتضح من الشكلين أن العلاقة بين الوزن وكل من العمر والطول علاقة خطية.

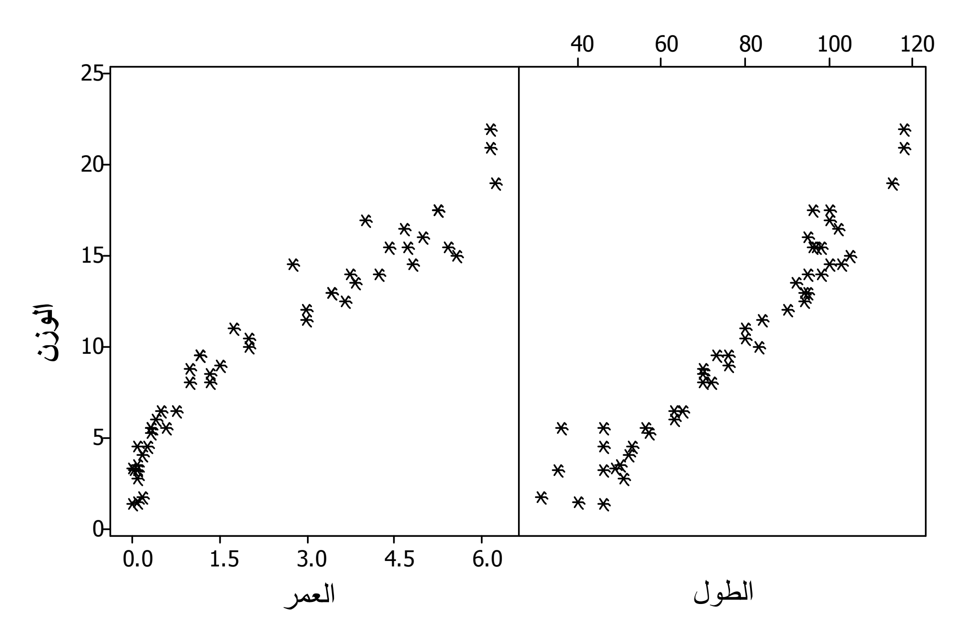
جدول رقم (٣-٢): أوزان وأعمار وأطوال عينة عشوائية قوامها ٥٠ طفلاً من منطقة عسير

رقم المشاهدة	وزن الطفل (كجم)	عمر الطفل (سنة)	طول الطفل (سم)
1	11.50	3.00	84
2	16.00	5.00	95
3	6.50	0.50	65
4	17.00	4.00	100
5	8.50	1.33	70
6	8.80	1.00	70
7	22.00	6.17	118
8	13.00	3.42	95
9	12.50	3.67	94
10	15.50	5.42	97
11	9.50	1.17	76
12	15.50	4.42	96
13	9.50	1.17	73
14	14.50	2.75	100
15	19.00	6.25	115
16	9.00	1.50	76
17	14.00	4.25	98
18	10.50	2.00	80
19	6.00	0.42	63
20	15.00	5.58	105
21	13.00	3.42	94
22	21.00	6.17	118
23	12.00	3.00	90
24	17.50	5.25	100
25	5.50	0.33	56
26	5.30	0.33	57
27	6.50	0.75	63
28	13.50	3.83	92
29	4.50	0.25	53
30	15.50	4.75	98
31	16.50	4.67	102
32	11.00	1.75	80
33	17.50	5.25	96
34	14.55	4.83	103
35	10.00	2.00	83
36	4.00	0.17	52
37	3.50	0.08	50
38	8.00	1.00	70
39	8.00	1.33	72
40	14.00	3.75	95
41	1.75	0.17	31
42	3.20	0.08	46
43	5.55	0.33	46
44	2.75	0.08	51
45	1.35	0.01	46
46	5.50	0.58	36
47	4.50	0.08	46
48	3.25	0.02	35
49	3.30	0.00	49
50	1.40	0.08	40

المصدر: مستشفى أبها للنساء والولادة والمراكز الصحية (١٩٩٩م)



شكل رقم (٣-١): شكل انتشار ثلاثي الأبعاد لمتغيرات الوزن والطول والعمر



شكل رقم (٣-٢): شكلا انتشار متغير الوزن مع كل من الطول والعمر

للحصول على مقدرات معالم نموذج الانحدار، يتم وضع البيانات في قالب المصفوفات كما يلي:

$$\mathbf{y}=\begin{pmatrix}11.50\\16.00\\6.50\\.\\.\\.\\1.40\end{pmatrix},\quad \mathbf{x}=\begin{pmatrix}1&3.00&84.00\\1&5.00&95.00\\1&0.50&65.00\\.&.&.\\.&.&.\\.&.&.\\1&0.08&40.00\end{pmatrix}$$

حيث إن:

y متجه عمودي من الدرجة (50×1) يحتوي على بيانات المتغير التابع (أوزان الأطفال).

X مصفوفة البيانات من الدرجة (50×3)، يحتوي العمود الأول على قيمة الواحد الصحيح والثاني والثالث على قيم المتغيرين العمر (X_1) والطول (X_2) على التوالي. ولإيجاد مقدرات المربعات الصغرى نستخدم الصيغة (3.17) على النحو الذي يلي:

$$(\mathbf{x}^{T}\mathbf{x})=\begin{pmatrix}1.00&1.00&1.00&.&.&.&1.00\\3.00&5.00&0.50&.&.&.&0.08\\84.00&95.00&65.00&.&.&.&40\end{pmatrix}\begin{pmatrix}1.00&3.00&84.00\\1.00&5.00&95.00\\1.00&0.50&65.00\\.&.&.\\.&.&.\\.&.&.\\1.00&0.08&40.0\end{pmatrix}=\begin{pmatrix}50.0&117.4&3820.0\\117.36&491.7&11278\\3820.0&11277.5&320090.0\end{pmatrix}$$

ومن ثم يتم حساب معكوس المصفوفة $(\mathbf{x}^{T}\mathbf{x})$ كما يلي:

$$(\mathbf{x}^{T}\mathbf{x})^{-1}=\begin{pmatrix}0.789947&0.144239&-0.0145092\\0.144239&0.036936&-0.0030227\\-0.014509&-0.003023&0.0002828\end{pmatrix}$$

وحساب المتجه $(\mathbf{x}^{T}\mathbf{y})$

$$(\mathbf{x}^{T}\mathbf{y})=\begin{pmatrix}1 & 1 & 1 & . & . & . & 1\\ 3 & 5 & 0.5 & . & . & . & 0.08\\ 84 & 95 & 65 & . & . & . & 40\end{pmatrix}\begin{pmatrix}11.50\\ 16.00\\ 6.50\\ .\\ .\\ .\\ 1.4\end{pmatrix}=\begin{pmatrix}507.7\\ 1739.2\\ 45081.8\end{pmatrix}$$

وبضرب $(\mathbf{x}^{T}\mathbf{x})^{-1}$ في $(\mathbf{x}^{T}\mathbf{y})$ نحصل على تقديرات المربعات الصغرى التالية:

$$\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\begin{pmatrix}0.789948 & 0.144239 & -0.014509\\ 0.144239 & 0.036936 & -0.003023\\ -0.014509 & -0.003023 & 0.000283\end{pmatrix}\begin{pmatrix}507.7\\ 1739.2\\ 45081.8\end{pmatrix}=\begin{pmatrix}-2.1818689\\ 1.20080469\\ 0.1245725\end{pmatrix}$$

وعليه فإن نموذج انحدار وزن الطفل على عمره وطوله يأخذ الشكل التالي:

$$\hat{y}= -2.18190 + 1.2008x_1 + 0.12457x_2$$

ويمكن تفسير المعامل الثابت على أنه يمثل القيمة المقدرة للمتغير التابع عندما تكون قيم المتغيرات المفسرة مساوية للصفر. وفي الواقع نجد أن هذا التفسير ليس صحيحاً في كل الحالات. فباتباع هذا التفسير نجد أن الوزن المقدر يكون سالباً (-٢,١٨ كيلوجرام) عندما يكون عمر الطفل وطوله يساويان الصفر!. ولكن هل يوجد طفل عمره صفر وطوله صفر؟!. كما يجب ملاحظة أن مشاهدات العينة لا تحتوي على قيم صفرية لكل من متغيري العمر والطول (انظر جدول رقم (٣-٢)). حيث تُراوح أعمار الأطفال بين يوم وست سنوات ونصف السنة والأطوال بين (٣١) سم و(١١٨) سم. وكما سبق ذكره يتعين على الباحث استخدام نموذج الانحدار للتنبؤ في نطاق قيم مشاهدات المتغيرات المستقلة التي استُخدمت في بناء النموذج. ففي هذا المثال يستخدم المعامل الثابت مع المعاملين الآخرين $\hat{\beta}_1$ و $\hat{\beta}_2$ لحساب القيم الموفقة. أما معامل الانحدار الجزئي (١,٢) فيعني أن زيادة عمر الطفل بسنة واحدة تصحبها زيادة في وزنه بمقدار (١,٢) كيلوجرام بافتراض ثبات الطول. وكذلك إن الزيادة في طول الطفل بواحد سنتيمتر تؤدي إلى زيادة في وزن الطفل بحوالي (١٢٥) جراماً بافتراض ثبات العمر.

٣-٧ خواص مقدرات المربعات الصغرى:

٣-٧-١ عدم التحيز (Unbiasedness):

يعني عدم التحيز أن القيمة المتوقعة لكل عنصر من عناصر المتجه $\hat{\boldsymbol{\beta}}$ تساوي العنصر المقابل في متجه المعالم الحقيقة ($\boldsymbol{\beta}$)، أي:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

$$= \begin{pmatrix} E(\hat{\beta}_0) \\ E(\hat{\beta}_1) \\ E(\hat{\beta}_2) \\ . \\ . \\ . \\ E(\hat{\beta}_p) \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{pmatrix} \tag{3-18}$$

البرهان:

نعلم أن معادلة نموذج الانحدار بصيغة المصفوفات (3.4) هي:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

وأن المعادلة الأساسية لمقدرات المربعات الصغرى (3.17) هي:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

وبوضع المعادلة (3.4) في المعادلة (3.17) نحصل على:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} \end{aligned}$$

وبأخذ التوقع (E) لطرفي المعادلة نحصل على:

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\boldsymbol{\varepsilon})$$

وبما أن القيمة المتوقعة لحد الخطأ حسب الاشتراط الخامس تساوي صفر ($E(\boldsymbol{\varepsilon}) = \mathbf{0}$) فإن:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \tag{3-19}$$

أي أن مقدرات المربعات الصغرى الاعتيادية غير متحيزة.

٣-٧-٢ الخطية (Linearity):

نعلم أن المعادلة الأساسية لمقدرات المربعات الصغرى (3.17) هي:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

وبما أن $\mathbf{(X^{T}X)^{-1}X^{T}}$ مصفوفة أرقام ثابتة، فإن $\widehat{\boldsymbol{\beta}}$ دالة خطية لـY ومن ثم فإن مقدرات المربعات الصغرى مقدرات خطية.

٣-٧-٣ خاصية أقل تباين (الكفاءة):

افترض أن ($\tilde{\boldsymbol{\beta}}$) أي مقدر خطي آخر لـ $\boldsymbol{\beta}$، أي:

$$\tilde{\boldsymbol{\beta}} = [(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T} + \mathbf{c}]\mathbf{y} \qquad (3\text{-}20)$$

حيث إن $\mathbf{c}$ مصفوفة من درجة $(p+1)\times n$ تتكون عناصرها من ثوابت.

وبما أن $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ فإن

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= [(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T} + \mathbf{c}]\left[\mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right] \\ &= (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\mathbf{x}\boldsymbol{\beta} + \mathbf{c}\mathbf{x}\boldsymbol{\beta} + (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon} + \mathbf{c}\boldsymbol{\varepsilon} \qquad (3\text{-}21) \\ &= \boldsymbol{\beta} + \mathbf{c}\mathbf{x}\boldsymbol{\beta} + (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon} + \mathbf{c}\boldsymbol{\varepsilon}\end{aligned}$$

ولكي يكون المتجه $\tilde{\boldsymbol{\beta}}$ مقدراً غير متحيز للمتجه $\boldsymbol{\beta}$ يجب أن تكون المصفوفة $\mathbf{cx}$ مصفوفة صفرية (Null matrix)، أي $\mathbf{cx} = \mathbf{0}$. وبالتالي يمكن إعادة كتابة المعادلة (3.21) كما يلي:

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon} + \mathbf{c}\boldsymbol{\varepsilon}$$

والآن يمكن تعريف تغاير-تباين المتجه $\tilde{\boldsymbol{\beta}}$ كما يلي:

$$\text{var}-\text{cov}(\tilde{\boldsymbol{\beta}}) = \text{E}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^{T}$$

$$\begin{aligned}\text{var-cov}(\tilde{\boldsymbol{\beta}}) &= \text{E}\left[(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon} + \mathbf{c}\boldsymbol{\varepsilon}\right]\left[(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon} + \mathbf{c}\boldsymbol{\varepsilon}\right]^{T} \\ &= \text{E}\left[(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{T}\mathbf{x}(\mathbf{x}^{T}\mathbf{x})^{-1} + \mathbf{c}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{T}\mathbf{x}(\mathbf{x}^{T}\mathbf{x})^{-1} + (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{T}\mathbf{c}^{T} + \mathbf{c}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{T}\mathbf{c}^{T}\right] \\ &= \sigma^{2}(\mathbf{x}^{T}\mathbf{x})^{-1} + \sigma^{2}\mathbf{c}\mathbf{x}(\mathbf{x}^{T}\mathbf{x})^{-1} + \sigma^{2}(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\mathbf{c}^{T} + \sigma^{2}\mathbf{c}\mathbf{c}^{T}\end{aligned}$$

وباستخدام $\mathbf{cx} = \mathbf{0}$ أو $\mathbf{x}^{T}\mathbf{c}^{T} = \mathbf{0}$ تصبح المعادلة أعلاه كالتالي:

$$\text{var-cov}(\tilde{\boldsymbol{\beta}}) = \sigma^{2}(\mathbf{x}^{T}\mathbf{x})^{-1} + \sigma^{2}\mathbf{c}\mathbf{c}^{T} \qquad (3\text{-}22)$$

وتوضح المعادلة (3.22) أن مصفوفة تباين-تغاير متجه المقدر الخطي غير المتحيز ($\tilde{\boldsymbol{\beta}}$) تساوي مصفوفة تباين-تغاير متجه المقدر الخطي غير المتحيز ($\widehat{\boldsymbol{\beta}}$) زائداً القيمة ($\sigma^{2}\mathbf{cc}^{T}$). ومن ثم فإن تباين كل عنصر من عناصر المتجه

($\tilde{\boldsymbol{\beta}}$) أكبر من أو يساوي تبـاين العنصر المقابـل في المتجـه ($\hat{\boldsymbol{\beta}}$). وبالتـالي لا توجـد مقـدرات تباينهـا أقـل مـن تبـاين مقدرات المربعات الصغرى، ولذلك فإن مقدرات المربعات الصغرى تتسم بالكفاءة لأنها خطية وغير متحيـزة ولهـا أقـل تباين.

٣-٨ خصائص البواقي:

البواقي هي القيم المقدرة لحد الخطـأ العشـوائي (ε_i) وهـي عبـارة عـن الفـروق بـين القـيم الفعليـة أو المُشـاهدة (Observed Values) والقيم المقدرة (Fitted values) المناظرة لها، أي:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}} \qquad (3\text{-}23)$$

وبما أن $\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}$ فإن

$$\mathbf{e} = \mathbf{y} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} = \mathbf{y} - \mathbf{H}\mathbf{y} \qquad (3\text{-}24)$$

حيث إن $\mathbf{H} = \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T$ مصفوفة من درجة $n \times n$. وتعرف المصفوفة $\mathbf{H}$ بمصفوفة القبعة (Hat matrix) لأنها تحول قيم المتغير التابع المشاهدة (Y) إلى القيم المقدرة ($\hat{y}$) التي تظهر فوقها علامة القبعة، أي:

$$\hat{\mathbf{y}} = \mathbf{x}\hat{\boldsymbol{\beta}} = \mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} = \mathbf{H}\mathbf{y} \qquad (3\text{-}25)$$

ومن خصائص مصفوفة القبعة أنها جامدة (Idempotent)، أي أن: $\mathbf{H}\mathbf{H}^T = \mathbf{H}^T\mathbf{H} = \mathbf{H}$

والآن يمكن كتابة متجه البواقي كما يلي:

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \mathbf{H}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y} \qquad (3\text{-}26) \\ &= \mathbf{M}\mathbf{y} \end{aligned}$$

حيث إن المصفوفة ($\mathbf{M} = (\mathbf{I} - \mathbf{H})$) هي أيضاً مصفوفة جامدة.

وفيما يلي نستعرض خصائص البواقي:

١) القيمة المتوقعة لأي عنصر من عناصر متجه البواقي (e) تساوي صفراً، أي:

$$\mathbf{E}(\mathbf{e}) = \mathbf{0} \qquad (3\text{-}27)$$

البرهان:

$$\begin{aligned}\mathbf{e} &= (\mathbf{I}-\mathbf{H})(\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon})\\ &= \mathbf{x}\boldsymbol{\beta}-\mathbf{H}\mathbf{x}\boldsymbol{\beta}+(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}\\ &= \mathbf{x}\boldsymbol{\beta}-\mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{x}\boldsymbol{\beta}+(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}\\ &= \mathbf{x}\boldsymbol{\beta}-\mathbf{x}\boldsymbol{\beta}+\mathbf{M}\boldsymbol{\varepsilon}\\ &= \mathbf{M}\boldsymbol{\varepsilon}\end{aligned}$$

وبأخذ التوقع لطرفي المعادلة نحصل على:

$$\mathrm{E}(\mathbf{e}) = \mathrm{E}(\mathbf{M}\boldsymbol{\varepsilon}) = \mathbf{M}\,\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$$

٢) استقلال البواقي عن المتغيرات المستقلة، أي:

$$\mathbf{x}^T\mathbf{e} = \mathbf{0} \qquad (3\text{-}28)$$

البرهان:

$$\mathbf{x}^T\mathbf{e} = \mathbf{x}^T\left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}\right) = \mathbf{x}^T\mathbf{y}-\mathbf{x}^T\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} = \mathbf{x}^T\mathbf{y}-\mathbf{x}^T\mathbf{y} = \mathbf{0}$$

٣) استقلال البواقي عن القيم المقدرة للمتغير التابع، أي:

$$\mathbf{y}^T\mathbf{e} = \mathbf{0} \qquad (3\text{-}29)$$

البرهان:

بما أن $\hat{\mathbf{y}} = \mathbf{x}\hat{\boldsymbol{\beta}}$ فإن

$$\hat{\mathbf{y}}^T\mathrm{e} = \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{e} == \hat{\boldsymbol{\beta}}^T\mathbf{0} = \mathbf{0}$$

ذلك لأن ($\mathbf{x}^T\mathbf{e} = \mathbf{0}$) من الخاصية الثانية.

٤) تباين متجه البواقي يساوي ($\sigma^2\mathbf{M}$)، أي:

$$\mathrm{V}(\mathbf{e}) = \sigma^2\mathbf{M} \qquad (3\text{-}30)$$

البرهان:

وبما أن $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ فإن

$$\mathrm{V}(\mathbf{e}) = \mathrm{E}\left[\mathbf{e}-\mathrm{E}(\mathbf{e})\right]\left[\mathbf{e}-\mathrm{E}(\mathbf{e})\right]^T = \mathrm{E}\left(\mathbf{e}\mathbf{e}^T\right)$$

وبما أن $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$ فإن

$$V(\mathbf{e}) = E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{M}^T) = \mathbf{M}\sigma^2\mathbf{I}\mathbf{M}^T = \sigma^2\mathbf{M}\mathbf{M}^T = \sigma^2\mathbf{M}$$

ذلك لأن المصفوفة $\mathbf{M}$ مصفوفة جامدة.

٥) مقدر التباين $\hat{\sigma}^2 = S^2 = \dfrac{\mathbf{e}^T\mathbf{e}}{n-p-1}$ مقدر غير متحيز لـ σ^2

البرهان:

نعلم أن $\mathbf{e}^T\mathbf{e} = \sum_{i=1}^{n} e_i^2$، وكذلك مجموع مربعات البواقي ($\sum_{i=1}^{n} e_i^2$) يساوي مجموع العناصر القطرية (Trace) للمصفوفة ($\mathbf{e}\mathbf{e}^T$) حيث

$$\mathbf{e}\mathbf{e}^T = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_n \end{pmatrix} \begin{pmatrix} e_1 & e_2 & e_3 & . & . & . & e_n \end{pmatrix} = \begin{pmatrix} e_1^2 & e_1e_2 & e_1e_3 & . & . & . & e_1e_n \\ e_2e_1 & e_2^2 & e_2e_3 & . & . & . & e_2e_n \\ e_3e_1 & e_3e_2 & e_3^2 & . & . & . & e_3e_n \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ e_ne_1 & e_ne_2 & e_ne_3 & . & . & . & e_n^2 \end{pmatrix}$$

أي أن:

$$\mathbf{e}^T\mathbf{e} = \text{trace}\left(\mathbf{e}\mathbf{e}^T\right)$$

إذن

$$E\left(\mathbf{e}^T\mathbf{e}\right) = E\left(\text{trace}\left(\mathbf{e}\mathbf{e}^T\right)\right) = \text{trace}\, E\left(\mathbf{e}\mathbf{e}^T\right) = \text{trace}\, E\left(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{M}\right) = \text{trace}\left(\sigma^2\mathbf{M}\right)$$

وذلك لأن المصفوفة $\mathbf{M}$ مصفوفة جامدة، ومن ثم

$$\begin{aligned} E(\mathbf{e}^T\mathbf{e}) &= \text{trace}(\sigma^2\mathbf{M}) \\ &= \sigma^2\text{trace}(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \\ &= \sigma^2[\text{trace}\mathbf{I} - \text{trace}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}] \\ &= \sigma^2(\text{trace}\ \mathbf{I}_{nxn} - \text{trace}\ \mathbf{I}_{(p+1)(p+1)}) \end{aligned}$$

وبما أن مصفوفة الوحدة ($\mathbf{I}$) يتكون عناصر قطرها من الواحد الصحيح، فإن:

$$E(\mathbf{e}^{T}\mathbf{e}) = \sigma^{2}(n-p-1)$$

ومن ثم فإن مقدر التباين هو:

$$\hat{\sigma}^{2} = \frac{\mathbf{e}^{T}\mathbf{e}}{n-p-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} \qquad (3\text{-}31)$$

مقدر غير متحيز لـ σ^2

٦) مجموع مربعات البواقي هو: $\mathbf{e}^{T}\mathbf{e} = \mathbf{y}^{T}\mathbf{y} - \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y}$

البرهان:

$$\begin{aligned}
\mathbf{e}^{T}\mathbf{e} &= \left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}\right)^{T}\left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}\right) \\
&= \mathbf{y}^{T}\mathbf{y} - \mathbf{y}^{T}\mathbf{x}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y} + \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{x}\hat{\boldsymbol{\beta}} \\
&= \mathbf{y}^{T}\mathbf{y} - \mathbf{y}^{T}\mathbf{x}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y} + \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\mathbf{y} \\
&= \mathbf{y}^{T}\mathbf{y} - \mathbf{y}^{T}\mathbf{x}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y} + \hat{\boldsymbol{\beta}}^{T}\mathbf{I}\mathbf{x}^{T}\mathbf{y}
\end{aligned}$$

وبما إن $(\mathbf{y}^{T}\mathbf{x}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y})$ فإن

$$\mathbf{e}^{T}\mathbf{e} = \mathbf{y}^{T}\mathbf{y} - \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y}$$

وباستخدام هذه المعادلة يمكن إعادة كتابة معادلة مقدر التباين على النحو التالي:

$$S^{2} = \frac{\mathbf{e}^{T}\mathbf{e}}{n-p-1} = \frac{\mathbf{y}^{T}\mathbf{y} - \hat{\boldsymbol{\beta}}^{T}\mathbf{x}^{T}\mathbf{y}}{n-p-1} \qquad (3\text{-}32)$$

٣-٩ مصفوفة تباين-تغاير معاملات الانحدار:

يُعرف تباين-تغاير متجه معاملات الانحدار المقدرة ($\hat{\boldsymbol{\beta}}$) كالتالي:

$$\text{var-cov}\left(\hat{\boldsymbol{\beta}}\right) = E\left\{\left[\hat{\boldsymbol{\beta}} - E\left(\hat{\boldsymbol{\beta}}\right)\right]\left[\hat{\boldsymbol{\beta}} - E\left(\hat{\boldsymbol{\beta}}\right)\right]^{T}\right\} \qquad (3\text{-}33)$$

وباستخدام شكل المصفوفات يمكن كتابة المعادلة (3.33) على النحو التالي:

$$\text{var}-\text{cov}(\hat{\boldsymbol\beta}) = \begin{pmatrix} \text{var}(\hat\beta_1) & \text{cov}(\hat\beta_1\hat\beta_2) & \text{cov}(\hat\beta_1\hat\beta_3) & . & . & . & \text{cov}(\hat\beta_1\hat\beta_p) \\ \text{cov}(\hat\beta_2\hat\beta_1) & \text{var}(\hat\beta_2) & \text{cov}(\hat\beta_2\hat\beta_3) & . & . & . & \text{cov}(\hat\beta_2\hat\beta_p) \\ \text{cov}(\hat\beta_3\hat\beta_1) & \text{cov}(\hat\beta_3\hat\beta_2) & \text{var}(\hat\beta_3) & . & . & . & \text{cov}(\hat\beta_3\hat\beta_p) \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ \text{cov}(\hat\beta_p\hat\beta_1) & \text{cov}(\hat\beta_p\hat\beta_2) & \text{cov}(\hat\beta_p\hat\beta_3) & . & . & . & \text{var}(\hat\beta_p) \end{pmatrix}$$

نعلم أن:

$$\hat{\boldsymbol\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$
$$\mathbf{y} = \mathbf{x}\boldsymbol\beta + \boldsymbol\varepsilon$$

إذن

$$\hat{\boldsymbol\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{x}\boldsymbol\beta + \boldsymbol\varepsilon)$$
$$\hat{\boldsymbol\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}\boldsymbol\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol\varepsilon$$
$$\hat{\boldsymbol\beta} = \mathbf{I}\boldsymbol\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol\varepsilon$$
$$\hat{\boldsymbol\beta} = \boldsymbol\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol\varepsilon$$

وبطرح ($\boldsymbol\beta$) من طرفي المعادلة نحصل على:

$$\hat{\boldsymbol\beta} - \boldsymbol\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol\varepsilon$$

ومن معادلة تباين-تغاير معاملات الانحدار نجد أن:

$$\text{var-cov}\left(\hat{\boldsymbol\beta}\right) = \text{E}\left\{\left[\hat{\boldsymbol\beta} - \text{E}\left(\hat{\boldsymbol\beta}\right)\right]\left[\hat{\boldsymbol\beta} - \text{E}\left(\hat{\boldsymbol\beta}\right)\right]^T\right\}$$
$$\text{var-cov}\left(\hat{\boldsymbol\beta}\right) = \text{E}\left\{\left[\hat{\boldsymbol\beta} - \boldsymbol\beta\right]\left[\hat{\boldsymbol\beta} - \boldsymbol\beta\right]^T\right\}$$
$$\text{var-cov}\left(\hat{\boldsymbol\beta}\right) = \text{E}\left\{\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol\varepsilon\right]\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol\varepsilon\right]^T\right\}$$
$$\text{var-cov}(\hat{\boldsymbol\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}$$
$$\text{var-cov}(\hat{\boldsymbol\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}$$
$$\text{var-cov}(\hat{\boldsymbol\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

وبما أن التباين مجهول يستخدم مقدره (S^2) وبذلك تصبح معادلة مقدر تباين وتغاير مقدرات المربعات الصغرى هو:

$$S^2\left(\hat{\boldsymbol{\beta}}\right) = S^2\left(\mathbf{x}^T\mathbf{x}\right)^{-1} \qquad (3.34)$$

٣-١٠ مثال:

من المثال السابق، احسب مصفوفة التباين-التغاير لمعاملات نموذج وزن الطفل ومن ثم أوجد تقديرات الأخطاء المعيارية لهذه المعاملات؟

الحل:

يتم أولاً حساب التباين (S^2) كما يلي:

$$S^2 = \frac{\mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}}{n-p-1}$$

وبما أن

$$\mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y} = (6653.3853 - (-2.1819\ 1.2008\ 0.12457)\begin{pmatrix} 507.7 \\ 1739.2 \\ 45081.8 \end{pmatrix}$$

$$= 6653.3853 - 6596.6948$$

$$= 56.6904$$

وأن

$$n\text{-}p\text{-}1=50\text{-}2\text{-}1=47$$

إذًا مقدر التباين هو:

$$S^2 = \tfrac{56.6904}{47} = 1.20618$$

وبضرب S^2 في المصفوفة $\left(\mathbf{x}^T\mathbf{x}\right)^{-1}$ التي حسابها في المثال السابق نحصل على مصفوفة تباين-تغاير مقدرات معالم نموذج الانحدار ($\hat{\boldsymbol{\beta}}$) كما يلي:

$$S^2\left(\hat{\boldsymbol{\beta}}\right) = S^2\left(\mathbf{x}^T\mathbf{x}\right)^{-1} = 1.20618\begin{pmatrix} 0.78948 & 0.144239 & -0.014509 \\ 0.144239 & 0.036936 & -0.003023 \\ -0.014509 & -0.003023 & 0.00283 \end{pmatrix} = \begin{pmatrix} 0.95282 & 0.173978 & -0.017501 \\ 0.173978 & 0.044551 & -0.003646 \\ -0.017501 & -0.003646 & 0.000341 \end{pmatrix}$$

وبأخذ الجذر التربيعي للعناصر القطرية في المصفوفة نحصل على الأخطاء المعيارية لمعاملات نموذج الانحدار الثلاثة كما يلي:

$$s.e(\widehat{\beta}_0) = \sqrt{S^2(\widehat{\beta}_0)} = \sqrt{0.95282} = 0.976125$$
$$s.e(\widehat{\beta}_1) = \sqrt{S^2(\widehat{\beta}_1)} = \sqrt{0.044551} = 0.21107$$
$$s.e(\widehat{\beta}_2) = \sqrt{S^2(\widehat{\beta}_2)} = \sqrt{0.000341} = 0.0184683$$

٣-١١ الاستدلال الإحصائي:

٣-١١-١ التقدير بفترة لمعاملات الانحدار الجزئية:

باستيفاء اشتراط التوزيع الطبيعي لحد الخطأ العشوائي نجد أن متجه معاملات الانحدار ($\widehat{\boldsymbol{\beta}}$) يتبع التوزيع الطبيعي بوسط حسابي يساوي المتجه ($\boldsymbol{\beta}$) وباين يساوي المصفوفة ($\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$)، أي:

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta},\ \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$$

ويعني هذا أن كل عنصر من عناصر المتجه ($\widehat{\boldsymbol{\beta}}$)، $\widehat{\beta}_k$ مثلاً، يتبع التوزيع الطبيعي بوسط حسابي يساوي العنصر المقابل في متجه المعالم ($\boldsymbol{\beta}$)، β_k، وتباين ($\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}_{kk}$)، أي:

$$\widehat{\beta}_k \sim N(\beta_k, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}_{kk}), \qquad k = 0,1,2,\ldots,p$$

حيث إن $(\mathbf{X}^T\mathbf{X})^{-1}_{kk}$ يمثل العنصر القطري رقم k من المصفوفة $(\mathbf{X}^T\mathbf{X})^{-1}$. وبالطريقة المعتادة يمكن حساب القيمة المعيارية Z حيث

$$Z = \frac{\widehat{\beta}_k - \beta_k}{\sqrt{\sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}_{kk}}} \sim N(0,1)$$

تتبع التوزيع الطبيعي المعياري بوسط حسابي يساوي الصفر وانحراف معياري يساوي الواحد الصحيح. وبما أن القيمة

$$\frac{(n-p-1)S^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

تتبع توزيع مربع كاي بدرجات حرية $(n-p-1)$، فإن إحصاء T حيث

$$T = \frac{(\hat{\beta}_k - \beta_k)/\sigma\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{kk}}}{\sqrt{S^2/\sigma^2}} \sim t_{n-p-1} \qquad (3-35)$$

تتبع توزيع t بدرجات حرية (n-p-1). وباستخدام هذه الإحصاءة يمكن اشتقاق فترة الثقة لـ β_k كما يلي:

$$\Pr[-t_{\alpha/2,n-p-1} \le \frac{\hat{\beta}_k - \beta_k}{s.e\left(\hat{\beta}_k\right)} \le t_{\alpha/2,n-p-1}] = 1-\alpha$$

وبإعادة ترتيب المعادلة أعلاه تصبح:

$$\Pr\left[\hat{\beta}_k - t_{\alpha/2,\,n-p-1}\,.\,s.e\left(\hat{\beta}_k\right) \le \beta_k \le \hat{\beta}_k + t_{\alpha/2,\,n-p-1}\,.\,s.e\left(\hat{\beta}_k\right)\right] = 1-\alpha$$

وبالتالي يمكن كتابة فترة الثقة $100(1-\alpha)$ للمعامل β_k كما يلي:

$$\hat{\beta}_k \pm t_{\alpha/2,n-p-1}.s.e.\left(\hat{\beta}_k\right) \qquad (3-36)$$

حيث إن:

$\hat{\beta}_k =$ معامل الانحدار رقم k، و k=0,1,2,...,p

$s.e.\left(\hat{\beta}_k\right) =$ مقدر الخطأ المعياري للمعامل $\hat{\beta}_k$

$t_{\alpha/2,n-p-1} =$ القيمة الحرجة التي يتم استخراجها من جدول توزيع t بدرجات حرية (n-p-1) ومستوى معنوية (α).

٣-١١-٢ اختبارات المعنوية لمعاملات الانحدار الجزئية:

تستخدم الإحصاءة (T) حسب المعادلة (3.35) لاختبار ما إذا كانت المعلمة β_k تأخذ قيمة معينة، β_k^* والفرض المراد اختباره في هذه الحالة هو:

فرض العدم ($H_0: \beta_k = \beta_k^*$) مقابل الفرض البديل ($H_1: \beta_k \ne \beta_k^*$)

حيث إن (β_k^*) قيمة يفترضها الباحث. وتستخدم الإحصاءة (T) بصفة أساسية لاختبار الفرض الصفري ($H_0 = \beta_k = 0$) ذلك لمعرفة مدى إسهام المتغير (X_k) في تفسير التغير في المتغير التابع. ويكون إحصاء الاختبار في هذه الحالة هي:

$$T = \frac{\hat{\beta}_k - 0}{s.e(\hat{\beta}_k)} \sim t_{n-p-1} \qquad (3-37)$$

فإذا كانت قيمة T المطلقة أكبر من قيمة توزيع t عند درجات حرية (n-p-1) ومستوى معنوية محدد نرفض فرض العدم ونقبل الفرض البديل وهو أن قيمة المعلمة (β_k) تختلف عن الصفر، وهذا يعني أن المتغير المستقل (X_k) يؤثر على المتغير التابع أو يسهم في تفسير تباين المتغير التابع. أما إذا كانت قيمة T المطلقة أقل من قيمة توزيع t الجدولية، يعني أننا ليس لدينا دليل كاف لرفض فرض العدم وبالتالي فإن المتغير المستقل (X_k) ليس له تأثير على المتغير التابع أو لا يسهم بمستوى معنوي في تفسير تباين المتغير التابع.

مثال:

احسب فترة ثقة (٩٥%) للمعلمتين β_1 و β_2 لنموذج انحدار وزن الطفل؟

الحل:

لإيجاد فترتي الثقة المناظرة لمعلمتي الانحدار β_1 و β_2 يتم حساب قيمتي الخطأ المعياري المناظرة لكل من $\widehat{\beta}_1$ و $\widehat{\beta}_2$ وهما: $s.e.\left(\widehat{\beta}_1\right)$ و $s.e.\left(\widehat{\beta}_2\right)$ على التوالي، وحساب قيمة توزيع t عند درجات حرية (٤٧) ومستوى معنوية (٠,٠٢٥).

ومن المثال السابق نجد أن قيمتي الانحراف المعياري لـ $\widehat{\beta}_1$ و $\widehat{\beta}_2$ وهما على التوالي:

$$s.e.\left(\widehat{\beta}_1\right) = 0.21107$$
$$s.e\left(\widehat{\beta}_2\right) = 0.0184683$$

وحيث إن

$$t_{\alpha/2,n-p-1} = t_{0.025,47} = 2.01174$$

فإن فترة ثقة (٩٥%) لمعامل الانحدار β_1 هي:

$$1.2008 \pm 2.001174 \times 0.21107$$
$$0.776 < \beta_1 < 1.625$$

وهذا معناه أننا واثقون بدرجة ٩٥% أن قيمة المعلمة الحقيقية (β_1) تقع في المدى ما بين ٠,٧٧٦ و١,٦٢٥. وكذلك نجد فترة ثقة (٩٥%) للمعلمة (β_2) هي:

$$0.12457 \pm 2.01174 \times 0.0184683$$
$$0.0874 < \beta_2 < 0.1617$$

وهذا يعني فترة ثقة (٩٥%) للمعلمة (β_2) تمتد ما بين ٠,٠٨٧ و٠,١٦٢.

مثال:

من مثال نموذج انحدار الطفل، اختبر معنوية/دلالة متغير العمر والطول كل على حده؟

أولاً اختبار معنوية المعلمة β_1:

الفرض المراد اختباره هو:

فرض العدم: $H_0{:}\beta_1 = 0$ مقابل الفرض البديل: $H_1{:}\beta_1 \neq 0$

وبما أن $\widehat{\beta}_1$ =1.2008 و $s.e\left(\widehat{\beta}_1\right) = 0.2111$ فإن:

$$T_1 = \frac{\widehat{\beta}_1}{s.e(\widehat{\beta}_1)} = \frac{1.2008}{0.2111} = 5.69$$

وبما أن قيمة T_1 أكبر من توزيع t عند درجات حرية (٤٧) ومستوى معنوية (٠,٠٥) ($t_{0.025,47}$ =2.012)، فإننا نرفض فرض العدم القائل بأن قيمة المعلمة (β_1) مساوية للصفر عند مستوى معنوية (٥%) ونستنتج أن متغير عمر الطفل يسهم في تفسير تباين الوزن.

ثانياً اختبار معنوية المعلمة β_2:

الفرض المراد اختباره هو:

فرض العدم: $H_1{:}\beta_2 = 0$ مقابل الفرض البديل: $H_1{:}\beta_2 \neq 0$

وبما أن $\widehat{\beta}_2$ =0.12457 و $s.e\left(\widehat{\beta}_2\right) = 0.01847$ فإن:

$$T_2 = \frac{\widehat{\beta}_2}{s.e(\widehat{\beta}_2)} = \frac{0.12457}{0.01847} = 6.74$$

وبما أن قيمة T_2 أكبر من توزيع t عند درجات حرية (٤٧) ومستوى معنوية (٥%) ($t_{0.025,47}$ =2.012)، فإننا نرفض فرض العدم القائل بأن قيمة المعلمة (β_2) مساوية للصفر عند مستوى معنوية (٥%) ونستنتج أن متغير طول الطفل يسهم في تفسير تباين الوزن.

٣-١١-٣ معامل التحديد (Coefficient of determination):

كما سبق شرحه في تحليل الانحدار الخطي البسيط، يقيس معامل التحديد نسبة التباين أو التغير في المتغير التابع (Y) التي تفسرها المتغيرات المستقلة $(X_1,X_2,\ldots,X_p)$، أي أنه يقيس نسبة التباين في Y التي يمكن تفسيرها بمعادلة الانحدار الموفقة.

ومن معادلة الانحدار الأساسية (الفصل الثاني) يمكن كتابة معادلة معامل التحديد على النحو التالي:

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \qquad (3\text{-}38)$$

وباستخدام رموز المصفوفات يمكن كتابة مكونات المعادلة الأساسية للانحدار كما يلي:

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2 = \mathbf{y}^T\mathbf{y} - n\bar{y}^2$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \mathbf{e}^T\mathbf{e} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}$$

$$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = TSS - RSS = \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y} - n\bar{y}^2$$

وباستخدام مكونات المعادلة الأساسية للانحدار في صيغ المصفوفات يمكن إعادة كتابة معادلة معامل التحديد (3.38) كما يلي:

$$R^2 = \frac{\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y} - n\bar{y}^2}{\mathbf{y}^T\mathbf{y} - n\bar{y}^2} \qquad (3\text{-}39)$$

ويلاحظ الآتي على معامل التحديد:

- تُراوح قيم معامل التحديد بين الصفر والواحد الصحيح، أي:

$$0 \le R^2 \le 1$$

- إذا كانت قيمة R^2 صغيرة فإن هذا يعني أن الجزء الأكبر من تباين Y يرجع إلى متغيرات لم تتضمن في نموذج الانحدار. أما إذا كانت قيمة R^2 قريبة من الواحد الصحيح فإن ذلك يعني أن الجزء الأكبر من تباين المتغير التابع قد تم تفسيره بواسطة المتغيرات المستقلة المضمنة في نموذج الانحدار.
- معامل التحديد R^2 دالة تزايدية لعدد المتغيرات المستقلة. فإضافة أي متغير مستقل لنموذج الانحدار تزيد من قيمة المعامل بغض النظر عن مساهمة هذا المتغير في تفسير تباين المتغير التابع. وللتخلص من هذا العيب يستخدم معامل آخر يعرف "بمعامل التحديد المعدل" (Adjusted Coefficient of Determination). ويتم حساب معامل التحديد المعدل ($\bar{R}^2$) حسب الصيغة التالية:

$$\bar{R}^2 = 1 - \frac{RSS/(n-p-1)}{TSS/(n-1)} = 1 - \frac{(TSS-ESS)/(n-p-1)}{TSS/(n-1)}$$
$$= 1 - \frac{(1-R^2)(n-1)}{n-p-1} \qquad (3\text{-}40)$$

كما يلاحظ الآتي على معامل التحديد المعدل:

- يأخذ قيماً أقل من قيم معامل التحديد (غير المعدل).
- يمكن أن يأخذ قيماً سالبة في حين نجد أن قيم معامل التحديد غير المعدل تكون دائماً موجبة.

٣-١١-٤ معامل الارتباط المتعدد (Multiple Correlation Coefficient):

يقيس معامل الارتباط المتعدد العلاقة الخطية بين المتغير التابع (Y) والمتغيرات المستقلة ($X_1, X_2, \ldots, X_p$). وهذا المعامل يمثل معامل الارتباط البسيط بين القيم الفعلية للمتغير التابع (Y_i) والقيم المقدرة ($\hat{Y}_i$). ويتم حساب معامل الارتباط المتعدد كما في حالة الارتباط الخطي البسيط حسب الصيغة التالية:

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\ \sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \qquad (3\text{-}41)$$

ويلاحظ الآتي على معامل الارتباط المتعدد:

- كلما اقتربت القيم المقدرة من قيم المشاهدات الفعلية للمتغير التابع ارتفعت قيمة معامل الارتباط المتعدد (R).
- يأخذ معامل الارتباط المتعدد قيماً غير سالبة وهو يختلف في هذه الصفة عن معامل الارتباط الخطي البسيط الذي يمكن أن يأخذ قيماً سالبة، أي: $0 \le R \le 1$
- معامل الارتباط المتعدد يساوي الجذر التربيعي الموجب لمعامل التحديد (R^2)، أي: $R = \left|\sqrt{R^2}\right|$

مثال:

من مثال نموذج وزن الطفل، احسب معامل التحديد ومعامل التحديد المعدل ومعامل الارتباط المتعدد.

الحل:

لإيجاد قيمة معامل التحديد يتم حساب القيم التالية:

$$\begin{aligned} ESS &= \hat{\boldsymbol{\beta}}^T \mathbf{x}^T \mathbf{y} - n\bar{y}^2 \\ &= 6596.7 - 5155.19 = 1441.51 \\ TSS &= \mathbf{y}^T \mathbf{y} - n\bar{y}^2 \\ &= 6653.39\text{-}5155.19 = 1498.20 \end{aligned}$$

وبالتعويض في معادلة معامل التحديد نحصل على:

$$R^2 = \frac{1441.51}{1498.20} = 0.96216$$

أي أن ٩٦,٢% من التباين أو التغير في أوزان الأطفال قد جرى تفسيره بواسطة عمر (X_1) وطول الطفل (X_2)؛ وهذا جزء كبير جداً يدعم وجود العلاقة الخطية.

أما معامل التحديد المعدل فإنه يتم حسابه على النحو التالي:

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{(n-1)}{(n-p-1)}(1-R^2) \\ &= 1\text{-} \frac{49}{47} \times (1-0.962161) = 0.960551 \end{aligned}$$

ويجري نفس تفسير معامل التحديد لمعامل التحديد المعدل، أي أن ٩٦% من التباين أو التغير في أوزان الأطفال يعزى لمتغيري العمر والطول.

كما يمكن حساب معامل الارتباط المتعدد بإيجاد الجذر التربيعي لمعامل التحديد كما يلي:

$$R = +\sqrt{R^2} = \sqrt{0.96216} = 0.9808975$$

يوضح هذا المعامل أن هناك علاقة خطية قوية تربط بين وزن الطفل وعمره وطوله.

٣-١١-٥ جدول تحليل التباين وإحصاء F:

لاختبار معنوية الانحدار ككل يستخدم جدول تحليل التباين للإجابة عن السؤال التالي:

هل يسهم جميع المتغيرات المستقلة المضمنة في نموذج الانحدار بمستوى معنوي في التنبؤ بقيم المتغير التابع؟ وبمعنى آخر هل تؤثر المتغيرات المستقلة كمجموعة تأثيراً ذا دلالة إحصائية على المتغير التابع أم لا؟. والفرض المراد اختباره في هذه الحالة هو:

فرض العدم: $H_0{:}\beta_0=\beta_1=\ldots=\beta_p=0$ مقابل **الفرض البديل:** ليست كل قيم المعالم مساوية للصفر.

وامتدادًا لأسلوب تحليل التباين في نموذج الانحدار الخطي البسيط يمكن كتابة جدول تحليل التباين الخاص بنموذج الانحدار الخطي المتعدد باستخدام صيغ المصفوفات على النحو التالي:

جدول (٣-٣): جدول تحليل التباين لنموذج الانحدار الخطي المتعدد

مصدر التغير Source of variation	درجات الحرية Degrees of Freedom	مجموع المربعات Sum of Squares	متوسط المربعات Mean Sum of Squares	قيمة F_0 F_0 Value
الانحدار	p	$\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}-n\bar{y}^2$	$\dfrac{\left(\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}-n\bar{y}^2\right)}{p}$	$\dfrac{\left(\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}-n\bar{y}^2\right)/p}{\left(\mathbf{y}^T\mathbf{y}-\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}\right)/(n-p-1)}$
البواقي	n-p-1	$\mathbf{y}^T\mathbf{y}-\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}$	$\dfrac{\mathbf{y}^T\mathbf{y}-\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}}{n-p-1}$	
المجموع	n-1	$\mathbf{y}^T\mathbf{y}-n\bar{y}^2$		

وباستيفاء اشتراط التوزيع الطبيعي لحد الخطأ نجد أن الإحصاءة F_o حيث

$$F_o=\frac{ESS/p}{RSS/(n-p-1)}=\frac{\left(\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}-n\bar{y}^2\right)/p}{\left(\mathbf{y}^T\mathbf{y}-\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}\right)/(n-p-1)}\sim F_{p,(n-p-1)} \qquad (3.42)$$

تتبع توزيع F بدرجتي حرية p و(n-p-1)، حيث n عدد المشاهدات وp عدد المتغيرات المستقلة. وبتحديد مستوى المعنوية يمكننا تحديد قاعدة القرار التالية:

-إذا كانت قيمة F_0 أكبر من قيمة توزيع F بدرجتي حرية p و(n-p-1)، نرفض فرض العدم القائل بتساوي جميع معالم النموذج للصفر، وبالتالي قبول الفرض البديل "ليس جميع المعالم تساوي قيمها الصفر" ونحكم بوجود علاقة خطية بين المتغير التابع (Y) وبعض أو جميع المتغيرات المستقلة $(X_1,X_2,\ldots,X_p)$.

- أما إذا كانت قيمة F_0 أقل من قيمة توزيع F عند درجتي حرية p و(n-p-1) فنقول إنه لا يوجد دليل كاف لرفض فرض العدم، وبالتالي نحكم بعدم وجود علاقة خطية بين المتغير التابع والمتغيرات المستقلة ذات دلالة إحصائية.

ويلاحظ الآتي على إحصائية F_0:

- من الممكن أن نحصل على قيمة كبيرة لـ F_0 دالة إحصائيّا وليس بين المعالم التي تم تقديرها ما هو معنوي إحصائيّا؛ يحدث ذلك عندما يكون هناك ارتباط خطي متعدد بين المتغيرات المستقلة (Multicollinearity) (انظر الفصل السابع).

- يمكن إجراء اختبار F_0 بدلالة معامل التحديد (R^2) كالتالي:

نعلم أن

$$F_0 = \frac{ESS}{RSS}\frac{(n-p-1)}{p}$$

أي أن:

$$F_0 = \frac{ESS}{(TSS-ESS)}.\frac{(n-p-1)}{p}$$

وبقسمة كل من بسط ومقام طرف المعادلة الأيمن على TSS نحصل على:

$$F_0 = \frac{ESS/TSS}{(1-ESS/TSS)}.\frac{(n-p-1)}{p}$$

$$F_0 = \frac{R^2}{(1-R^2)}.\frac{(n-p-1)}{p} \quad \sim \quad F_{p,(n-p-1)} \tag{3.43}$$

ويلاحظ من المعادلة (3.43) أنه إذا كانت قيمة R^2 تساوي صفراً فإن قيمة F تساوي صفراً أيضاً، وكلما زادت قيمة R^2 زادت قيمة F وعندما تقترب قيمة إلى الواحد الصحيح تقترب قيمة F إلى ما لا نهاية.

مثال:

من مثال نموذج انحدار وزن الطفل، هل يسهم المتغيران (العمر والطول) بمستوى معنوي في التنبؤ بوزن الطفل؟

الحل:

للإجابة عن هذا السؤال يتم صياغة فرض العدم التالية:

فرض العدم: $H_0: \beta_0 = \beta_1 = \beta_2 = 0$ مقابل الفرض البديل: ليست جميع قيم هذه المعالم مساوية للصفر.

ولإجراء هذا الاختبار تستخدم الإحصاءة F_0 حيث

$$F_o = \frac{\left(\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y} - n\bar{y}^2\right)/p}{\left(\mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y}\right)/\left(n-p-1\right)} \sim F_{p,(n-p-1)}$$

التي لها توزيع F بدرجتي حرية p و(n-p-1).

أو

$$F_0 = \frac{R^2}{(1-R^2)}.\frac{(n-p-1)}{p} \sim F_{p,(n-p-1)}$$

ولإيجاد قيمة الإحصاءة F_0 يتم حساب القيم التالية:

$$\hat{\boldsymbol{\beta}}^T \mathbf{x}^T \mathbf{y} = 6596.7$$
$$n\bar{y}^2 = 5155.19$$
$$\mathbf{y}^T \mathbf{y} = 6653.39$$
$$p = 2, \quad n = 50$$

إذن

$$F_0 = \frac{(6596.7 - 5155.19)}{(6653.39 - 6596.7)} \cdot \frac{47}{2} = 597.49$$

كما يمكن الحصول على نفس قيمة F_0 باستخدام المعادلة (3.44):

$$F_0 = \frac{R^2}{(1 - R^2)} \cdot \frac{(n - p - 1)}{p} = \frac{0.96216}{(1 - 0.96216)} \cdot \frac{47}{2} = 597.49$$

وبما أن قيمة الإحصاءة F_0 تزيد كثيراً على القيمة الحرجة ($F_{2,47,0.01}$=5.087)، نرفض فرض العدم عند مستوى معنوية أقل من (١%)، مما يدل على وجود علاقة خطية بين وزن الطفل وكل من عمره وطوله. وتوضع نتيجة اختبار دلالة الانحدار ككل في جدول تحليل التباين. ويتطلب جدول تحليل التباين حساب القيم التالية:

$$ESS = \hat{\boldsymbol{\beta}}^T \mathbf{x}^T \mathbf{y} - n\bar{y}^2$$
$$= 6596.7 - 5155.19 = 1441.51$$
$$RSS = \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{x}^T \mathbf{y}$$
$$= 6653.39 - 6596.70 = 56.70$$
$$TSS = \mathbf{y}^T \mathbf{y} - n\bar{y}^2$$
$$= 6653.39 - 5155.19 = 1498.20$$

جدول تحليل تباين نموذج انحدار وزن الطفل على عمره وطوله

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	قيمة F_0
الانحدار	٢	١٤٤١,٥١	٧٢٠,٧٥	٥٩٧,٤٩
البواقي	٤٧	٥٦,٧٠	١,٢١	
المجموع	٤٩	١٤٩٨,٢٠		

٣-١٢التقدير والتنبؤ

كما في تحليل الانحدار الخطي البسيط يستخدم نموذج الانحدار المقدر في:

أ) تقدير القيمة المتوسطة/ المتوقعة للمتغير التابع التي تقابل قيمًا معينة للمتغيرات المستقلة (x_0)

ب) التنبؤ بمشاهدة جديدة للمتغير التابع التي تقابل قيمًا معينة للمتغيرات (x_0).

٣-١٢-١ تقدير القيمة المتوسطة لـ Y:

يتم الحصول على تقدير القيمة المتوسطة لـ ($\hat{y}_0$) بالتعويض في نموذج الانحدار المقدر كما يلي:

$$\underset{1\times1}{\left(\hat{\mathbf{y}}_0 \middle| \mathbf{x}_0^T\right)} = \underset{1\times(p+1)}{\mathbf{x}_0^T} \underset{(p+1)\times1}{\hat{\boldsymbol{\beta}}} \qquad (3\text{-}44)$$

حيث إن:

$\mathbf{x}_0$ متجه صفي يحتوي على قيم المتغيرات المستقلة المراد عندها تقدير القيمة المتوسطة للمتغير التابع، أي أن:

$$\mathbf{x}_0^T = \left(1\ x_{10}\ x_{20}\ x_{30}\ \ldots\ x_{p0}\right)$$

$\hat{\boldsymbol{\beta}}$ متجه عمودي يحتوي على معالم نموذج الانحدار المقدرة، أي:

$$\hat{\boldsymbol{\beta}}^T = \left(\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2 \quad . \quad . \quad . \quad \hat{\beta}_p\right)$$

ويعتبر هذا التقدير أفضل تقدير خطي غير متحيز (Best Linear Unbiased Estimator) للقيمة المتوقعة، كما أن لهذا المقدر أقل تباين هو:

$$\begin{aligned}\text{var}(\hat{\mathbf{y}}_0) &= \text{var}(\mathbf{x}_0\hat{\boldsymbol{\beta}}) \\ &= \mathbf{x}_0^T \text{var-cov}(\hat{\boldsymbol{\beta}})\mathbf{x}_0 \\ &= \mathbf{x}_0^T \sigma^2 (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}_0\end{aligned}$$

و باستيفاء فرضية تبعية حد الخطأ العشوائي للتوزيع الطبيعي نجد أن $\hat{y}_0$ يتبع التوزيع الطبيعي بوسط حسابي وتباين هما: $E(\mathbf{x}_0^T\hat{\boldsymbol{\beta}}) = \mathbf{x}_0^T\boldsymbol{\beta}$ و $\sigma^2\mathbf{x}_0(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}_0^T$ على التوالي، أي أن:

$$\hat{\mathbf{y}}_0 \sim N(\mathbf{x}_0^T, \ \ \sigma^2\mathbf{x}_0^T(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}_0)$$

وبالطريقة المعتادة لتحويل متغير طبيعي إلى متغير طبيعي معياري يمكننا حساب القيمة Z حيث

$$Z = \frac{\hat{\mathbf{y}}_0 - \mathbf{x}_0^T\boldsymbol{\beta}}{\sigma\sqrt{\mathbf{x}_0^T(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}_0}} \quad \sim \quad N(0, 1)$$

يتبع التوزيع الطبيعي المعياري بوسط حسابي قدره صفر وانحراف معياري يساوي واحدًا صحيحًا. وبما أن $\frac{(n-p-1)S^2}{\sigma^2} \sim \chi^2$ تتبع توزيع مربع كاي χ^2 فإن القيمة T حيث:

$$T = \frac{\widehat{\mathbf{y}}_0 - \mathbf{x}_0^T \boldsymbol{\beta}}{S\sqrt{\mathbf{x}_0^T (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}_0}} \sim t_{n-p-1} \quad (3-45)$$

لها توزيع t بدرجات حرية (n-p-1). وباستخدام هذه المعادلة يمكن اشتقاق فترة الثقة للقيمة المتنبأ بها $\widehat{y}_0$ على النحو التالي:

$$\Pr(-t_{\alpha/2,n-p-1} \le \frac{\widehat{\mathbf{y}}_0 - \mathbf{x}_0^T \boldsymbol{\beta}}{S\sqrt{\mathbf{x}_0^T (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}_0}} \le t_{\alpha/2,n-p-1}) = 1-\alpha$$

وبإعادة ترتيب هذه المعادلة نحصل على فترة الثقة للقيمة المتوسطة للمتغير التابع كما يلي:

$$\mathbf{x}_0^T \widehat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-p-1} . S\sqrt{\mathbf{x}_0^T (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}_0} \quad (3-46)$$

مثال:

بالنسبة لنموذج انحدار وزن الطفل على عمره وطوله، يمكننا تقدير القيمة المتوسطة لوزن الطفل عندما يكون عمره ٦,٥ وطوله ١٢٠ سنتيمتراً وإيجاد فترة ثقة (٩٥%) لهذا التقدير.

الحل:

بما أن عمر وطول الطفل المراد عندهما التقدير بالقيمة المتوسطة لوزن الطفل يساويان ٦,٥ سنة و١٢٠ سم على التوالي، فإن متجه الصف $\mathbf{x}_0$ هو:

$$\mathbf{x}_0^T = \begin{pmatrix} 1 & 6.5 & 120 \end{pmatrix}$$

وعليه فإن تقدير القيمة المتوسطة هو:

$$\widehat{\mathbf{y}}_0 = \mathbf{x}_0^T \widehat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 6.5 & 120 \end{pmatrix} \begin{pmatrix} -2.1819 \\ 1.20080 \\ 0.12457 \end{pmatrix} = 20.5717kg$$

أي أن وزن الطفل يبلغ في المتوسط ٢٠,٦ كيلوجرامات، عندما يكون عمره ٦,٥ سنوات، وطوله ١٢٠ سنتيمتراً. ولحساب فترة ثقة ٩٥% يتم استخدام المعادلة (3.46) كما يلي:

$$\widehat{y}_0 \pm t_{\alpha/2,n-p-1} . s.e(\widehat{y}_0)$$

ولإيجاد الخطأ المعياري للقيمة المتنبأ بها يتم حساب مقدر تباين القيمة المتوسطة كما يلي:

$$V(\hat{\mathbf{y}}_0)=\mathbf{x}_0^T s^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$$

$$V(\hat{\mathbf{y}}_0)=\begin{pmatrix}1 & 6.5 & 120\end{pmatrix}\begin{pmatrix}0.95282 & 0.173978 & -0.017501\\ 0.173978 & 0.044551 & -0.003646\\ -0.017501 & -0.003646 & 0.000341\end{pmatrix}\begin{pmatrix}1\\ 6.5\\ 120\end{pmatrix}=0.120543$$

ومن جدول توزيع t ستودنت نجد أن $t_{0.025,47}=2.0117$ وعليه فإن فترة الثقة هي:

$$20.5717\pm(2.0117)\sqrt{0.120543}$$
$$(19.873\ \ ,\ \ 21.271)$$

وهذا يعني أن متوسط وزن الطفل عندما يكون عمره ٦,٥ سنوات وطوله ١٢٠ سنتيمتراً يُراوح بين ١٩,٩ كيلوجرامات حدًّا أدنى و٢١,٢٧ كيلوجرام حدًّا أعلى وذلك عند مستوى معنوية ٥% أو درجة ثقة ٩٥%.

٣-١٢-٢ التنبؤ بمشاهدة جديدة:

إن قيمة المشاهدة الجديدة للمتغير التابع (y_0) ترتبط بالقيمة المقدرة لها ($\hat{y}_0$) حسب المعادلة التالية:

$$y_0=\hat{y}_0+\varepsilon_0=\mathbf{x}_0^T\hat{\boldsymbol{\beta}}+\varepsilon_0$$

حيث إن ε_0 هو حد خطأ عشوائي جديد يتبع التوزيع الطبيعي بوسط حسابي يساوي الصفر وتباين قدره σ^2 وهو مستقل عن حد الخطأ (ε_0). وباستخدام اشتراط حد الخطأ العشوائي نجد أن:

$$\begin{aligned}E(y_0)&=E(\hat{y}_0+\varepsilon_0)\\ &=E(\mathbf{x}_0^T\hat{\boldsymbol{\beta}})+E(\varepsilon_0)\\ &=\mathbf{x}_0^T\hat{\boldsymbol{\beta}}+0=\mathbf{x}_0^T\hat{\boldsymbol{\beta}}\end{aligned}$$

وأن تباين (Y_0) هو:

$$\begin{aligned}\text{var}(y_0)&=\text{var}(\hat{y}_0+\varepsilon_0)\\ &=\text{var}(\hat{y}_0)+\text{var}(\varepsilon_0)+2\text{cov}(\hat{y}_0,\varepsilon_0)\\ &=\text{var}(\mathbf{x}_0^T\hat{\boldsymbol{\beta}})+\sigma^2+0\\ &=\sigma^2(\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0)+\sigma^2\\ &=\sigma^2[1+\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0)\end{aligned}$$

وبما أن قيمة σ^2 مجهولة فإنه يتم التعويض عنها بمقدر التباين s^2 في المعادلة لتصبح:

$$s^2(y_0)\ =\ s^2\left[1+\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0\right]$$

وباتباع نفس الطريقة التي اتُّبعت لحساب فترة ثقة القيمة المتوسطة لـ Y نجد أن فترة التنبؤ لمشاهدة جديدة (Y_0) هي:

$$\mathbf{x}_0^{T}\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-p-1}.s\sqrt{\left(1+\mathbf{x}_0^{T}(\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{x}_0\right)} \quad (3\text{-}47)$$

حيث إن $t_{\alpha/2,n-p-1}$ هي قيمة توزيع t بدرجات حرية (n-p-1) ومستوى معنوية (α). ويلاحظ أن فترة التنبؤ للمشاهدة الجديدة أكثر اتساعاً من فترة التنبؤ للقيمة المتوسطة لـY والذي يتمثل في الحد الزائد s^2.

مثال:

بالنسبة لنموذجِ انحدار وزن الطفل على عمره وطوله يمكننا التنبؤ بوزن الطفل عندما يكون عمره ٦,٥ سنوات وطوله ١٢٠ سنتيمترا وإيجاد فترة ثقة (٩٥%) للتنبؤ بمشاهدة جديدة.

الحل:

بالنسبة للقيمة المتنبأ بها هي نفس القيمة المتوسطة المتنبأ (Y_0) أي أن:

$$Y_0 = 20.5717$$

ولحساب فترة ثقة ٩٥% للتنبؤ بمشاهدة جديدة نحتاج أولاً لحساب الخطأ المعياري للقيمة المتنبأ بها ($s.e(Y_0)$):

$$s.e(y_0) = \sqrt{s^2\left(1+\mathbf{x}_0^{T}\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{x}_0\right)}$$

وبما أن قيمة $s^2\mathbf{x}_0^{T}(\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{x}_0 = 0.1205$ و$s^2 = 1.21$، فإن الخطأ المعياري لـ (Y_0) هو:

$$s.e(y_0) = \sqrt{1.21+0.1205} = 1.15347$$

وبالتالي يمكن حساب فترة الثقة (٩٥%) للتنبؤ بمشاهدة جديدة كما يلي:

$$20.5717 \pm 2.0117 \times 1.15347$$

أي:

$$(18.252 \quad , \quad 22.892)$$

أي أن الحد الأدنى لفترة الثقة هو (١٨,٢٥٢) كيلوجرام والحد الأعلى هو (٢٢,٨٩٢) كيلوجرام.

٣-١٣ مبدأ مجموع المربعات الإضافي (Extra Sum of Squares Principle):

مبدأ مجموع المربعات الإضافي هو أسلوب آخر لاختبار دلالة معاملات الانحدار. حيث يقيس أسلوب مجموع المربعات الإضافي الزيادة الحدية في مجموع مربعات الانحدار (ESS) عند إضافة متغير واحد أو أكثر على المتغيرات المستقلة الأخرى المضمنة في نموذج الانحدار. فمثلاً إذا كان لدينا ثلاثة متغيرات مستقلة (X_1, X_2, X_3)، فإذا قمنا ببناء

نموذج انحدار يضم المتغيرات الثلاثة ويسمى **بالنموذج الكامل** (Full model) ثم قمنا ببناء نموذج انحدار آخر يضم المتغيرين (X_1, X_2) فقط ويسمى **بالنموذج المخفض** (Reduced Model)، فإن الفرق بين مجموع مربعات انحدار النموذج الكامل ومجموع مربعات انحدار النموذج المخفض هو مقدار مجموع المربعات الإضافي الذي أسهم به المتغير X_3 عند ضمه للمتغيرين X_2 وX_1 أي هو الإسهام الخاص بالمتغير X_3 بعد استبعاد إسهام المتغيرين X_2 وX_1.

٣-١٣-١ اختبار معنوية إضافة متغير واحد على المتغيرات المستقلة الأخرى المضمنة في نموذج الانحدار:

لاختبار **فرض العدم** الذي نصه "إضافة المتغير Z لنموذج الانحدار الذي يحتوي أصلاً على المتغيرات المستقلة $(X_1,X_2,\ldots,X_p)$ لا تسهم بمستوى معنوي أو ذي دلالة إحصائية في التنبؤ بقيم المتغير التابع (Y)" مقابل **الفرض البديل** الذي نصه "إن إضافة المتغير Z للنموذج الذي يحتوي على المتغيرات $(X_1,X_2,\ldots,X_p)$ تسهم بمستوى معنوي في التنبؤ بقيم المتغير Y"، تستخدم إحصائية الاختبار التالية:

$$F_0(Z|X_1,X_2,\ldots,X_p) = \frac{SS(Z|X_1,X_2,\ldots,X_p)/1}{RSS(X_1,X_2,\ldots,X_p,Z)/(n-p-2)} \qquad (3\text{-}48)$$

حيث إن:

$SS(Z|X_1,X_2,\ldots,X_p)$ = مجموع المربعات الإضافي ويساوي مجموع مربعات الانحدار للنموذج الكامل الذي يضم جميع المتغيرات المستقلة بما في ذلك المتغير Z ناقصاً مجموع مربعات انحدار النموذج المخفض الذي يضم المتغيرات $(X_1,X_2,\ldots,X_p)$.

$RSS(X_1,X_2,\ldots,Xp, Z)$ = مجموع مربعات البواقي للنموذج الكامل.

n-p-2 = درجات الحرية المناظرة لمجموع مربعات بواقي النموذج الكامل.

وتتبع الإحصاءة F_0 توزيع F بدرجتي حرية 1 وn-p-2، والفرض المراد اختباره هنا هو:

فرض العدم: $H_0{:}\beta_Z = 0$ مقابل **الفرض البديل:** $H_1{:}\beta_Z \neq 0$

ويمكن الوصول إلى قرار بشأن معنوية β_Z على النحو التالي:

- إذا كانت قيمة F_0 أكبر من قيمة F الجدولية عند مستوى معنوية محدد، نرفض فرض العدم ونقبل الفرض البديل ومن ثم فإن قيمة المعلمة β_Z لا تساوي صفراَ ويمكن القول في هذه الحالة إن المتغير Z يسهم بصورة معنوية في تفسير تباين المتغير التابع Y.
- وأما إذا كانت قيمة F_0 أقل من قيمة F الجدولية فإننا نقبل فرض العدم ومن ثم فإن المتغير Z لا يسهم بمستوى معنوي في تفسير تباين المتغير التابع.

ويلاحظ أن اختبار F_0 يكافئ اختبار T لاختبار الفرض التالي:

فرض العدم: $H_0:\beta_Z = 0$ مقابل **الفرض البديل:** $H_1:\beta_Z \neq 0$

حيث إن β_Z معامل المتغير Z في معادلة الانحدار التالية:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \beta_Z z + \varepsilon$$

والإحصاءة المعادلة لإحصاء F_0 لاختبار فرض العدم هي:

$$|T| = \frac{\hat{\beta}_Z}{s.e\left(\hat{\beta}_Z\right)} \sim t_{n-p-2}$$

حيث $\hat{\beta}_Z$ مقدر المعلمة β_Z و $s.e\left(\hat{\beta}_Z\right)$ هو الخطأ المعياري لـ $\hat{\beta}_Z$.

فإذا كانت قيمة T المطلقة أكبر من القيمة الجدولية $t_{\alpha/2,n-p-2}$ نرفض فرض العدم وبالتالي يمكن القول إن المتغير Z يسهم بمستوى معنوي في تفسير تباين المتغير التابع Y. وأما إذا كانت قيمة T المطلقة أقل من القيمة الجدولية $t_{\alpha/2,n-p-2}$ فإننا نقبل فرض العدم ومن ثم نخلص إلى أن المتغير Z لا يسهم إسهاماً ذا دلالة إحصائية في تفسير تباين المتغير التابع.

٣-١٣-٢ اختبار F الجزئي المتعدد (Multiple-Partial F-test):

يُعَدُّ اختبارF الجزئي المتعدد امتداداً لاختبار F الجزئي الأحادي. حيث يتم في هذا الاختبار تقييم إسهام إضافة اثنين أو أكثر من المتغيرات المستقلة للمتغيرات الأخرى المضمنة أصلاً في نموذج الانحدار. أي أننا نود اختبار ما إذا كان إضافة K متغير مستقل $(Z_1,Z_2,\ldots,Z_k)$ تسهم بمستوى معنوي في التنبؤ بقيم المتغير التابع Y بوجود المتغيرات المستقلة $(X_1,X_2,\ldots,X_p)$. ولإجراء هذا الاختبار يتم بناء نموذج الانحدار الكامل التالي:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \beta^*_1 z_1 + \beta^*_2 z_2 + \ldots + \beta^*_k z_k + \varepsilon$$

وبناء نموذج الانحدار المخفض التالي:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$$

والفرض المراد اختباره هو:

فرض العدم: $H_0:\beta^*_1 = \beta^*_2 = \ldots = \beta^*_k = 0$ مقابل **الفرض البديل:** "ليس كل قيم هذه المعالم مساوية للصفر"

ولإجراء اختبار F الجزئي المتعدد نحسب أولاً مجموع المربعات الإضافي الناتج عن إضافة المتغيرات $(Z_1,Z_2,\ldots,Z_k)$ لنموذج الانحدار على النحو التالي:

مجموع المربعات الإضافي= مجموع مربعات انحدار النموذج الكامل - مجموع مربعات انحدار النموذج المخفض.

مجموع المربعات الإضافي= مجموع مربعات بواقي النموذج المخفض - مجموع مربعات بواقي النموذج الكامل.

أو

$$SS(Z_1,Z_2,\ldots,Z_k|\ X_1,X_2,\ldots,X_p)= ESS(Z_1,Z_2,\ldots,Z_k\ X_1,X_2,\ldots,X_p) - ESS(X_1,X_2,\ldots,X_p)$$

$$SS(Z_1,Z_2,\ldots,Z_k|\ X_1,X_2,\ldots,X_p)= RSS(X_1,X_2,\ldots,X_p) - RSS(Z_1,Z_2,\ldots,Z_k\ X_1,X_2,\ldots,X_p)$$

ولإجراء الاختبار يتم حساب الإحصاءة F_0 حيث

$$F_0(Z_1,Z_2,...,Z_k|X_1,X_2,...,X_p)=\frac{SS(Z_1,Z_2,...,Z_k|X_1,X_2,...,X_p)/k}{RSS(Z_1,Z_2,...,Z_k,X_1,X_2,...,X_p)/(n-p-k-1)} \qquad (3\text{-}49)$$

وتتبع هذه الإحصاءة توزيع F بدرجتي حرية k و(n-p-k-1). ويمكن كتابة الإحصاءة (3.49) على النحو التالي:

$$F_0(Z_1,Z_2,...,Z_k|X_1,X_2,...,X_p)=\frac{[ESS(full)-ESS(reduced)]/k}{RSS(full)/(n-p-k-1)}$$

حيث إن:

ESS(full) = مجموع مربعات الانحدار للنموذج الكامل.

ESS(reduced) = مجموع مربعات الانحدار للنموذج المخفض.

RSS(full) = مجموع مربعات بواقي النموذج الكامل.

k = عدد المتغيرات المضافة المراد اختبار مساهمتها في تفسير المتغير التابع.

p = عدد المتغيرات المستقلة المضمنة أصلاً في النموذج.

n = عدد المشاهدات.

وبقسمة بسط ومقام الطرف الأيمن المعادلة (3.49) على مجموع المربعات الكلي (TSS) نحصل على الصيغة التالية:

$$F_0(Z_1,Z_2,...,Z_k|X_1,X_2,...,X_p)=\frac{[R^2(full)-R^2(reduced)]/k}{(1\text{-}R^2(full))/(n-p-k-1)} \qquad (3\text{-}50)$$

حيث إن:

$R^2(full)$= معامل تحديد النموذج الكامل.

$R^2(reduced)$ = معامل تحديد النموذج المخفض.

k = عدد المتغيرات المضافة المراد اختبار مساهمتها في تفسير المتغير التابع.

p = عدد المتغيرات المستقلة المضمنة أصلاً في النموذج.

n = عدد المشاهدات.

مثال:

بالنسبة لنموذج انحدار وزن الطفل، اختبر معنوية إضافة متغير العمر لنموذج الانحدار الذي يضم الطول وكذلك اختبر معنوية إضافة متغير الطول لنموذج الانحدار الذي يضم العمر باستخدام مبدأ مجموع المربعات الإضافي؟

الحل:

- **اختبار معنوية إضافة متغير العمر:**

لاختبار معنوية إضافة متغير العمر نحتاج إلى بناء نموذج الانحدار الكامل الذي يضم المتغيرين (العمر والطول) ونموذج الانحدار المخفض الذي يضم متغير الطول فقط كما يلي:

نموذج الانحدار الكامل: $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + e$

نموذج الانحدار المخفض: $y = \hat{\beta}_0 + \hat{\beta}_2 x_2 + e$

ومن إحصاءات النموذجين نحصل على القيم التالية:

مجموع مربعات انحدار النموذج الكامل $ESS(X_1,X_2)$ = ١٤٤١,٥٠٣٢٩٧

مجموع مربعات بواقي النموذج الكامل $RSS(X_1,X_2)$ = ٥٦,٦٩٥٩٠٢٧

مجموع مربعات انحدار النموذج المخفض $ESS(X_2)$ = ١٤٠٢,٤٦٤٢

والآن باستخدام مبدأ مجموع المربعات الإضافي نجد أن:

$$F_0(X_1|X_2) = \frac{[ESS(X_1,X_2) - ESS(X_2)]/1}{RSS(X_1,X_2)/(n-p-1)}$$

$$= \frac{(1441.503297\text{-}1402.4642)/1}{56.6959027/(50\text{-}2\text{-}1)} = 32.3628$$

وحيث إن قيمة إحصاء F_0 الجزئي أكبر بكثير من قيمة توزيع F بدرجتي حرية ١ و٤٧ $(F_{0.01,1,47}=7.21)$ فإننا نرفض فرض العدم عند مستوى معنوية (١%)، ويمكننا القول إن إضافة متغير العمر تسهم بمستوى معنوي في تفسير تباين المتغير التابع.

- **اختبار معنوية إضافة متغير الطول:**

لاختبار معنوية إضافة متغير الطول نتبع نفس طريقة اختبار متغير العمر. حيث يتم بناء نموذج كامل يضم المتغيرين (الطول والعمر) ونموذج مخفض يضم متغير العمر فقط:

ومن إحصاءات النموذجين نتحصل على القيم التالية:

مجموع مربعات انحدار النموذج الكامل $ESS(X_1,X_2)$ = ١٤٤١,٥٠٣٢٩٧

مجموع مربعات بواقي النموذج الكامل $RSS(X_1,X_2)$ = ٥٦,٦٩٥٩٠٢٧

مجموع مربعات انحدار النموذج المخفض $ESS(X_1)$ = ١٣٨٦,٦٢٤٨

والآن باستخدام مبدأ مجموع المربعات الإضافي نجد أن:

وحيث إن قيمة إحصاء F_0 الجزئي أكبر بكثير من قيمة توزيع F بدرجتي حرية ١ و٤٧ $(F_{0.01,1,47}=7.21)$

$$F_0(X_2\,|\,X_1)=\frac{[ESS(X_1,X_2)-ESS(X_1)]/\,1}{RSS(X_1,X_2)/(n\text{-}p\text{-}1)}$$
$$=\frac{\left(1441.503297-1386.6248\right)/\,1}{56.6959027\,/\,(50\text{-}2\text{-}1)}=45.493$$

فإننا نرفض فرض العدم عند مستوى معنوية (١%) ويمكننا القول إن إضافة متغير الطول تسهم بمستوى معنوي في تفسير تباين المتغير التابع. وتوضح النتائج التي تم استعراضها أن المتغيرين (العمر والطول) يؤثران بمستوى معنوي على وزن الطفل، إلا أن متغير الطول حين يستخدم لوحده للتنبؤ بقيم وزن الطفل أفضل من متغير العمر حين يستخدم لوحده؛ وذلك لأن مجموع مربعات الانحدار الإضافي لمتغير الطول (٥٤,٨٧٩) أكبر من مجموع مربعات الانحدار الإضافي الذي يسهم متغير العمر (٣٩,٠٣٩). ويوضح الجدول التالي جدول تحليل التباين لنماذج الانحدار الكامل والمخفضة.

جدول تحليل التباين لنماذج الانحدار الكامل والمخفضة

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	قيمة F
انحدار (X_1,X_2)	٢	١٤٤١,٥٠	٧٢٠,٧٥	٥٩٧,٤٩
انحدار (X_1)	١	١٣٨٦,٦٣	١٣٨٦,٦٣	٥٩٦,٥٤
انحدار (X_2)	١	١٤٠٢,٤٦	١٤٠٢,٤٦	٧٠٣,١٧
انحدار (X_1) بعد استبعاد (X_2)	١	٣٩,٠٤	٣٩,٠٤	٣٢,٣٦
انحدار (X_2) بعد استبعاد (X_1)	١	٥٤,٨٨	٥٤,٨٨	٤٥,٤٩
بواقي النموذج الكامل	٤٧	٥٦,٦٩	١,٢٠٦٣	
المجموع	٤٩	١٤٩٨,٢٠		

٣-١٤ معامل الارتباط الجزئي (Partial Correlation Coefficient):

٣-١٤-١ مقدمة:

يقيس معامل الارتباط الجزئي قوة واتجاه العلاقة الخطية بين متغيرين بعد عزل أو استبعاد أثر المتغيرات الأخرى. وإن الفرق بين معامل الارتباط الخطي البسيط ومعامل الارتباط الجزئي، هو أن الأول يقيس قوة واتجاه العلاقة بين متغيرين ضمن تأثيرات المتغيرات الأخرى، في حين يقيس الثاني قوة واتجاه العلاقة بين متغيرين بعد استبعاد تأثير المتغيرات الأخرى. فمثلاً إذا كان لدينا ثلاثة متغيرات (X_2, X_1, Y) فمن الممكن قياس الارتباط بين أي اثنين منهم مع عزل أثر الثالث باستخدام معامل الارتباط الجزئي. ويرمز لمعامل الارتباط الجزئي بين المتغير Y و X_1 مثلاً بعد استبعاد أثر المتغير X_2 بالتالي:

$$\rho_{YX_1.X_2}$$

ويسمى معامل الارتباط في هذه الحالة بمعامل الارتباط الجزئي من المرتبة الأولى (First-Order partial)، حيث يساوي مرتبة المعامل عدد المتغيرات المستبعد أثرها التي تعرف بالمتغيرات الضابطة (Control variables). ويسمى معامل الارتباط الخطي البسيط بمعامل الارتباط من المرتبة صفر (Zero-Order Correlation) لعدم وجود متغيرات تم استبعاد تأثيرها. وبصورة عامة يأخذ معامل الارتباط الجزئي من المرتبة (p) (p^{th} order) الصيغة التالية:

$$\rho_{YX_1.X_2,X_3,\dots,X_p,X_{p+1}}$$

٣-١٤-٢ معامل الارتباط الجزئي من المرتبة الأولى:

يتم حساب معامل الارتباط الجزئي من المرتبة الأولى بين Y وX_1 بعد استبعاد أثر المتغير X_2 والذي يرمز له بـ $r_{YX_1.X_2}$ حسب الصيغة التالية:

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{(1-r^2_{YX_2})(1-r^2_{X_1X_2})}} \qquad (3\text{-}51)$$

حيث إن r_{YX1}، r_{YX2}، وr_{X1X2} هي معاملات الارتباط البسيط بين أزواج المتغيرات (Y وX_1) و(Y وX_2) و(X_1 وX_2) على التوالي. وبالمثل يمكن حساب معامل الارتباط الجزئي بين Y و X_2 كما يلي:

$$r_{YX_2.X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{(1-r^2_{YX_1})(1-r^2_{X_1X_2})}}$$

ولاختبار فرض العدم القائل بأن معامل الارتباط الجزئي للمجتمع يساوي الصفر $H_0: \rho_{YX_1.X_2} = 0$ في مقابل الفرض البديل القائل بأنه يختلف عن الصفر $H_0: \rho_{YX_1.X_2} \neq 0$، يستخدم إحصائية الاختبار التالية:

$$T = \frac{r_{YX_1.X_2}\sqrt{n-3}}{\sqrt{(1-r^2_{YX_1.X_2})}} \sim t_{(n-3)} \qquad (3\text{-}52)$$

وتتبع إحصائية الاختبار T توزيع t بدرجات حرية $(n-3)$. فإذا كانت قيمة معامل الارتباط الجزئي كبيرة بدرجة كافية فإن قيمة الاختبار T المطلقة ستكون أيضاً كبيرة وهذا يدل على وجود علاقة قوية بين Y وX_1 بعد عزل أثر المتغير X_2 وبالتالي يُرفض فرض العدم القائل بعدم وجود علاقة ارتباط معنوية بين المتغير YوX_1 في المجتمع بعد تثبيت أثر المتغير X_2.

٣-١٤-٣ معامل الارتباط الجزئي من المرتبة الثانية:

يعتبر معامل الارتباط الجزئي من المرتبة الثانية امتداداً لمعامل الارتباط الجزئي من المرتبة الأولى. حيث يأخذ معامل الارتباط الجزئي بين Y وX_1 بعد استبعاد أثر المتغيرين X_2 وX_3 الصيغة التالية:

$$r_{YX_1.X_2X_3} = \frac{r_{YX_1.X_2} - r_{YX_3.X_2} r_{X_1X_3.X_2}}{\sqrt{(1-r^2_{YX_3.X_2})(1-r^2_{X_1X_3.X_2})}} = \frac{r_{YX_1.X_3} - r_{YX_2.X_3} r_{X_1X_2.X_3}}{\sqrt{(1-r^2_{YX_2.X_3})(1-r^2_{X_1X_2.X_3})}} \qquad (3\text{-}53)$$

ويتم إختبار معنوية معامل الارتباط الجزئي بين Y وX_1 بعد تثبيت أثر المتغيرين X_2 وX_3 بنفس الطريقة التي سبق شرحها وذلك باستخدام الإحصاءة T حيث

$$T=\frac{r_{YX_1.X_2X_3}\sqrt{n-4}}{\sqrt{1-r^2_{YX_1.X_2X_3}}} \sim t_{n-4} \qquad (3\text{-}54)$$

وبصورة عامة يتم اختبار دلالة معامل الارتباط الجزئي من المرتبة p باستخدام الصيغة التالية:

$$T=r_{YX_1.X_2,X_3,...,X_p,X_{p+1}}\frac{\sqrt{n-p-2}}{\sqrt{1-r^2_{YX_1.X_2,X_3,...,X_p,X_{p+1}}}} \sim t_{n-p-2} \qquad (3\text{-}55)$$

٣-١٤-٤ معاملات التحديد الجزئية (Coefficients of Partial Determination):

يقيس معامل التحديد الجزئي المساهمة الحدية لمتغير مستقل واحد (X) في تفسير التباين أو التغير في المتغير التابع عندما تكون المتغيرات الأخرى مضمنة في النموذج. وفي حالة نموذج انحدار خطي يضم متغيرين مستقلين حيث يأخذ النموذج الصيغة التالية:

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\varepsilon_i$$

يقيس مجموع مربعات البواقي $RSS(X_2)$ التباين أو التغير في Y عندما يضم النموذج المتغير X_2 فقط، في حين يقيس مجموع مربعات البواقي $RSS(X_1,X_2)$ التباين أو التغير في Y عندما يضم النموذج المتغيرين X_1 وX_2 ومن ثم فإن نسبة الانخفاض في تباين أو تغير المتغير التابع التي تعزى لإضافة X_1 للنموذج الذي يحتوي على X_2 هي:

$$\frac{RSS(X_2)-RSS(X_1,X_2)}{RSS(X_2)}$$

تعرف هذه النسبة بمعامل التحديد الجزئي بين Y وX_1 عندما يكون X_2 في النموذج ويرمز له بـ $r^2_{YX1.X2}$، أي أن:

$$r^2_{YX_1.X_2}=\frac{RSS(X_2)-RSS(X_1X_2)}{RSS(X_2)} \qquad (3\text{-}56)$$

وبالمثل يمكن حساب معامل التحديد الجزئي بين Y وX_2 عندما يكون X_1 في النموذج كما يلي:

$$r^2_{YX_2.X_1}=\frac{RSS(X_1)-RSS(X_1X_2)}{RSS(X_1)}$$

وبصورة عامة يمكن إيجاد معامل التحديد الجزئي بين المتغير Y والمتغير المستقل X_1 بعد استبعاد أثر المتغيرات $(X_2, X_3,...,X_p)$ كما يلي:

$$r^2_{YX_1.X_2,X_3,...,X_p}=\frac{RSS(X_2,X_3,...,X_p)-RSS(X_1,X_2,...,X_p)}{RSS(X_2,X_3,...,X_p)}$$

أما دلالة معامل التحديد الجزئي فيتم اختباره باستخدام اختبار F الجزئي المتعدد كما سبق شرحه في الجزء (٣-١٣-٢)، حيث تأخذ إحصائية الاختبار الصيغة التالية:

$$F(X_1.X_2,X_3,...,X_p)=\frac{RSS(X_2,X_3,...,X_p)-RSS(X_1,X_2,....X_p)}{RSS(X_1,X_2,....X_p)/(n-p-1)} \sim F_{1,(n-p-1)}$$

مثال:

من بيانات نموذج وزن الطفل على العمر والطول، احسب معاملات الارتباط البسيط ومن ثم معاملات الارتباط الجزئي والتحديد الجزئي بين كل من متغير الوزن والعمر وبين متغير الوزن والطول؟

الحل:

توضح المصفوفة التالية معاملات الارتباط البسيط بين المتغيرات الثلاثة (الوزن، الطول والعمر).

مصفوفة معاملات الارتباط بين متغيرات وزن الطفل، العمر والطول

المتغير	الوزن	العمر	الطول
الوزن	١,٠٠٠٠	٠,٩٦٢٠	٠,٩٦٧٥
العمر	٠,٩٦٢٠	١,٠٠٠٠	٠,٩٣٥٣
الطول	٠,٩٦٧٥	٠,٩٣٥٣	١,٠٠٠٠

- معامل الارتباط الخطي الجزئي بين وزن الطفل والعمر بعد عزل تأثير متغير الطول:

$$r_{\text{weight age.height}} = \frac{r_{\text{weight age}} - r_{\text{weight height}}\ r_{\text{age height}}}{\sqrt{(1-r^2_{\text{weight height}})(1-r^2_{\text{age height}})}}$$

$$= \frac{0.962 - 0.9675\times 0.9353}{\sqrt{(1-0.9675^2)(1-0.9353^2)}} = 0.638579$$

أي أن معامل الارتباط الجزئي بين وزن الطفل وعمره بعد عزل تأثير طول الطفل قد بلغ ٠,٦٤.

ولإجراء اختبار معنوية معامل الارتباط تستخدم إحصائية الاختبار التالية:

$$T=\frac{r_{YX_1X_2}\sqrt{n-3}}{\sqrt{(1-r^2_{YX_1/X_2})}}=\frac{0.638579\sqrt{47}}{\sqrt{(1-0.638579^2)}}=5.689$$

وبما أن قيمة توزيع t عند درجات حرية (٤٧) ومستوى معنوية (٠,٠٥) هي ٢,٠١٤ $(t_{47,0.05}=2.014)$ أقل بكثير من قيمة T المحسوبة، فإننا نرفض فرض العدم (H_0) في مقابل الفرض البديل (H_1)، أي أن بيانات العينة تدل بدرجة كافية على أن معامل الارتباط الخطي الجزئي بين وزن الطفل وعمره يختلف عن الصفر.

معامل الارتباط الجزئي بين وزن الطفل وطوله بعد استبعاد أثر متغير العمر:

$$r_{\text{weight height.age}} = \frac{r_{\text{weight height}} - r_{\text{weight age}}\ r_{\text{age height}}}{\sqrt{(1-r^2_{\text{weight age}})(1-r^2_{\text{age height}})}}$$

$$= \frac{0.9675 - 0.9620 \times 0.9353}{\sqrt{(1-0.9620^2)(1-0.9353^2)}} = 0.701324$$

ويتم إجراء اختبار معنوية معامل الارتباط بنفس الطريقة السابقة، أي:

$$T = \frac{0.701324\sqrt{47}}{\sqrt{(1-0.701324^2)}} = 6.745$$

وبما أن قيمة توزيع t عند درجات حرية (٤٧) ومستوى دلالة (٠,٠٥) أقل بكثير من قيمة إحصاء T المحسوبة، فإننا نرفض فرض العدم ونحكم على أن معامل الارتباط الخطي الجزئي بين وزن الطفل وطوله يختلف عن الصفر. وتشير النتائج إلى أن العلاقة الخطية بين الوزن والطول بعد عزل تأثير العمر أقوى من العلاقة الخطية بين وزن الطفل والعمر بعد عزل تأثير الطول.

معامل التحديد الجزئي بين متغيري الوزن والعمر بعد استبعاد متغير الطول:

يتطلب حساب معامل التحديد الجزئي في هذه الحالة إجراء نموذجين يضم أحدهما متغير الطول والآخر يضم متغيري العمر والطول ومن ثم يتم حساب مجموع مربعات البواقي لكل نموذج. وبتعويض قيمتي مجموع مربعات البواقي في معادلة معامل التحديد الجزئي نحصل على:

$$r^2_{\text{weight.age.height}} = \frac{RSS(height) - RSS(height.age)}{RSS(height)}$$

$$= \frac{95.735 - 56.696}{95.735} = 0.4078$$

وبالمثل يمكن حساب معامل التحديد الجزئي بين متغير الوزن والطول بعد عزل متغير العمر كما يلي:

$$r^2_{\text{weight height.age}} = \frac{RSS(age) - RSS(height.age)}{RSS(age)}$$

$$= \frac{111.574 - 56.696}{111.574} = 0.492$$

ويلاحظ هنا أن معامل التحديد الجزئي ما هو إلا عبارة عن مربع معامل الارتباط الجزئي. ففي هذا المثال نجد أن معامل التحديد الجزئي بين الوزن والطول بعد عزل تأثير متغير العمر يساوي مربع معامل الارتباط الجزئي، أي:

$$0.701324^2=0.4918556$$

وتوضح هذه النتائج أن إضافة متغير الطول لنموذج الانحدار الذي يضم متغير العمر تؤدي إلى انخفاض مجموع مربعات البواقي بنسبة ٤٩%، في حين تؤدي إضافة متغير العمر لنموذج الانحدار الذي يضم متغير الطول إلى انخفاض مجموع مربعات البواقي بنسبة ٤١%، مما يدل على أن مساهمة متغير الطول في تفسير تباين وزن الطفل -المتغير التابع- أكبر من مساهمة العمر.

٣-١٥ نموذج الانحدار المعياري (Standardized Regression Model):

٣-١٥-١ مقدمة:

قد نرغب أحياناً في معرفة الأهمية النسبية للمتغيرات المستقلة المضمنة في نموذج الانحدار. ولكن عندما تختلف وحدات قياس المتغيرات المستقلة يصبح من الصعب مقارنة قيم معاملات نموذج الانحدار لهذه المتغيرات لمعرفة أيهما أكثر تأثيراً على المتغير التابع. فمثلاً في نموذج انحدار وزن الطفل نجد أن متغير العمر مقاس بالسنوات في حين تم قياس الطول بالسنتيمتر وبالتالي حصلنا على أحجام معاملات مختلفة. وكذلك من المتوقع أن تتغير قيم معاملي المتغيرين إذا تغيرت وحدات قياسهما؛ فمن من الممكن أن يكون العمر بالأيام أو بالشهور وكذلك يمكن قياس الطول بالبوصات أو الأمتار. ولذلك نجد أن حجم المعامل لا يزود الباحث بمقياس يوضح أهمية المعامل في النموذج. ولكن من الممكن التخلص من تأثير وحدات القياس المختلفة للمتغيرات على أحجام معاملاتها ذلك بتحويلها إلى متغيرات معيارية (Standardized variables) ومن ثم يتم تقدير معالم نموذج الانحدار باستخدام طريقة المربعات الاعتيادية.

٣-١٥-٢ طريقة تقدير معالم نموذج الانحدار المعياري:

يأخذ نموذج الانحدار الخطي الصيغة التالية:

$$y_i = \hat{\beta}_0+\hat{\beta}_1 x_{1i}+\hat{\beta}_2 x_{2i}+\ldots+\hat{\beta}_p x_{pi}+e_i \qquad (3\text{-}57)$$

وبجمع وقسمة طرفي هذه المعادلة على (n) نحصل على:

$$\bar{y}_i = \hat{\beta}_0+\hat{\beta}_1\bar{x}_1+\hat{\beta}_2\bar{x}_2+\ldots+\hat{\beta}_p\bar{x}_p \qquad (3\text{-}58)$$

وبطرح المعادلة (3.58) من (3.57) نحصل على:

$$y_i-\bar{y}_i=\hat{\beta}_1\left(x_{1i}-\bar{x}_1\right)+\hat{\beta}_2\left(x_{2i}-\bar{x}_2\right)+\ldots+\hat{\beta}_p\left(x_{pi}-\bar{x}_p\right)+e_i \qquad (3\text{-}59)$$

ولتحويل المتغير التابع إلى متغير معياري نقسم طرفي المعادلة على الانحراف المعياري للمتغير التابع $\left(S_y=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}\right)$.

$$\left(\frac{y_i-\bar{y}_i}{s_y}\right)=\hat{\beta}_1\frac{(x_{1i}-\bar{x}_1)}{s_y}+\hat{\beta}_2\frac{(x_{2i}-\bar{x}_2)}{s_y}+\ldots+\hat{\beta}_p\frac{(x_{pi}-\bar{x}_p)}{s_y}+\frac{e_i}{s_y} \qquad (3\text{-}60)$$

وبضرب وقسمة كل حد من حدود الطرف الأيمن من المعادلة (3.60) على الانحراف المعياري S_{x_j} للمتغير المستقل X_j نحصل على:

$$\left(\frac{y_i-\bar{y}_i}{s_y}\right)=\hat{\beta}_1\frac{s_{x_1}}{s_{x_1}}\frac{(x_{1i}-\bar{x}_1)}{s_y}+\hat{\beta}_2\frac{s_{x_2}}{s_{x_2}}\frac{(x_{2i}-\bar{x}_2)}{s_y}+\ldots+\hat{\beta}_p\frac{s_{x_p}}{s_{x_p}}\frac{(x_{pi}-\bar{x}_p)}{s_y}+\frac{e_i}{s_y}$$

أو

$$y_i^*=\hat{\beta}_1^*x_{1i}^*+\hat{\beta}_2^*x_{2i}^*+\ldots+\hat{\beta}_p^*x_{pi}^*+e_i^* \qquad (3\text{-}61)$$

حيث إن: $y_i^*=\frac{(y_i-\bar{y})}{s_y}$ المتغير التابع المعياري، و $x_{ij}^*=\frac{(x_{ji}-\bar{x}_j)}{S_{xj}}$ المتغير المستقل المعياري رقم j ‏(j=1,2,…,p)،

و $\hat{\beta}_j^*=\hat{\beta}_j\frac{s_{x_j}}{s_y}$ معامل الانحدار الجزئي للمتغير المستقل المعياري رقم **j**، و $e_i^*=\frac{e_i}{s_y}$ الباقي وله وسط حسابي يساوي صفراً وانحراف معياري يختلف عن الواحد الصحيح.

ويمثل المعامل المعياري ($\hat{\beta}_j^*$) التغير في القيمة المتوقعة للمتغير التابع (y^*) بوحدات الانحراف المعياري الناتج عن تغير (x_j^*) بانحراف معياري واحد بافتراض ثبات قيم المتغيرات المستقلة المعيارية الأخرى. وهكذا يمكن تفسير بقية معاملات الانحدار المعيارية الجزئية الأخرى. وبالتالي من الممكن قياس الأثر النسبي للمتغيرات المستقلة، فيمكننا القول بأن أثر المتغير X_j أكبر (أقل) من أثر المتغير X_h إذا كانت قيمة المعامل المعياري الجزئي للمتغير X_j أكبر (أقل) من قيمة المعامل المعياري الجزئي للمتغير X_h.

٣-١٥-٣ مثال:

بالنسبة لنموذج وزن الطفل يمكننا استخدام الانحدار المعياري لمعرف أي المتغيرين (العمر، الطول) أكثر تأثيراً على وزن الطفل كما يلي:

بما أن قيم الانحراف المعياري لمتغيرات الوزن والعمر والطول على التوالي هي:

$$s_Y = 5.52951$$
$$s_{x1} = 2.10062$$
$$s_{x2} = 24.00765$$

فإن معاملي الانحدار المعياري هما:

$$\hat{\beta}_1^* = \hat{\beta}_1 \frac{s_{x1}}{s_Y} = 1.2008 \times \frac{2.10062}{5.52952} = 0.4561$$
$$\hat{\beta}_2^* = \hat{\beta}_2 \frac{s_{x2}}{s_Y} = 0.12457 \times \frac{24.00765}{5.52952} = 0.5408$$

أي أن نموذج الانحدار المعياري هو:

$$\hat{y}^* = 0.4561x_1^* + 0.5408x_2^*$$

وعلى الرغم من أن حجم معامل الطول في نموذج الانحدار العادي أقل بكثير من حجم معامل عمر الطفل، إلا أن حجم معامل متغير الطول في الانحدار المعياري أكبر من معامل متغير العمر المعياري. وهذا يعني أن زيادة الطول بانحراف معياري واحد تؤدي إلى زيادة أكبر في وزن الطفل (بوحدات الانحراف المعياري) عندما يكون العمر ثابتاً من زيادة العمر بانحراف معياري واحد عندما يكون الطول ثابتاً.

٣-١٥-٤ملاحظات:

- يجب أن لا تُستخدم المعاملات المعيارية لعمل مقارنات أثر نفس المتغير المستقل لعينات مسحوبة من مجتمعات مختلفة وذلك لاختلاف تباين المتغير من عينة لأخرى.
- يتضمن مخرجات بعض حزم البرامج الإحصائية الجاهزة كبرنامج SPSS معاملات الانحدار المعيارية $\hat{\beta}_k^*$ التي تسمى بمعاملات بيتا (Standardized coefficients Beta)، كما يوضح الإطار رقم (١).
- يجب توخي الحذر عند مقارنة الانحدارات المعيارية في حالة وجود ارتباط بين المتغيرات المستقلة لتأثر قيمها بذلك.

إطار رقم (١): نتائج انحدار وزن الطفل على الوزن والعمر- مخرجات برنامج SPSS

معاملات الانحدار غير المعيارية

معاملات الانحدار المعيارية

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	-2.182-	.976		-2.235-	.030
1	age.years	1.201	.211	.456	5.689	.000
	length.cm	.125	.018	.541	6.745	.000

a. Dependent Variable: weight.kg

٣-١٦ تحليل البواقي (Residuals Analysis):

يعد تحليل البواقي من أهم مراحل بناء نموذج الانحدار الخطي والذي يهدف إلى تقييم مدى ملاءمة نموذج الانحدار الذي تم بناؤه. وفي هذا الجزء نستعرض بعض الفحوص التشخيصية باستخدام تحليل البواقي للتأكد من استيفاء النموذج الموفق للاشتراطات التي تمت مناقشتها في الجزء (٣-٣). والجدير بالذكر أن هذه الفحوص تستخدم لنماذج الانحدار الخطي البسيط والانحدار الخطي المتعدد على حد سواء.

٣-١٦-١ البواقي والبواقي المعيارية:

البواقي (Residuals):

البواقي هي الفروق بين القيم الفعلية أو المُشاهدة (Observed Values) والقيم المقدرة (Fitted values) المناظرة لها، أي:

$$e_i = y_i - \hat{y}_i \qquad i = 1,2,...,n$$

حيث إن e_i الباقي رقم i، y_i القيمة المشاهدة للمتغير التابع رقم i و $\hat{y}_i$ القيمة الموفقة المناظرة لها.

البواقي المعيارية (Standardized Residuals):

يأخذ تباين البواقي الصيغة التالية (الجزء (٣-٨)):

$$V(e_i) = (1 - h_{ii})\, S^2$$

حيث إن h_{ii} هو العنصر القطري رقم i من مصفوفة القبعة $\mathbf{H}=\mathbf{X}\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T}$. ويلاحظ من هذه المعادلة أن تباين البواقي يعتمد على مصفوفة البيانات $\mathbf{X}$ وتختلف قيمه تبعاً لقيمة العنصر القطري (h_{ii}). وباستخدام الطريقة المعتادة يتم تحويل البواقي إلى متغير معياري يعرف بالبواقي المعيارية (e_i') على النحو التالي:

$$e_i' = \frac{e_i - \bar{e}}{\sqrt{v(e_i)}}$$

وبما أن $\bar{e}=0$ فإن الباقي المعياري رقم (i) هو

$$e_i' = \frac{e_i}{\left\{\left(1-h_{ii}\right)S^2\right\}^{1/2}} \qquad (3\text{-}62)$$

٣-١٦-٢ فحص النموذج (Model Diagnostic):

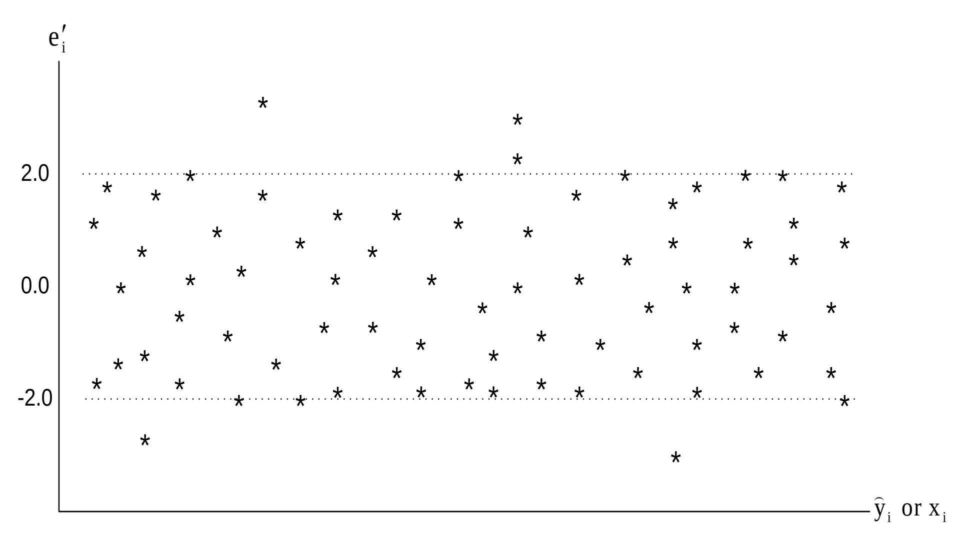
يستخدم تحليل البواقي للكشف عن خمسة أنواع من الانحرافات المحتملة هي:

- عدم خطية دالة الانحدار.
- عدم ثبات تباين حدود الخطأ.
- عدم استقلالية حدود الخطأ في حالة أن بيانات المتغير التابع والمتغيرات المستقلة مشاهدات سلسلة زمنية.
- عدم تبعية حدود الخطأ للتوزيع الطبيعي.
- الكشف عن المشاهدات الشاذة/الخارجة (Outlying observations).

وفيما يلي نستعرض بعض الطرق البيانية باستخدام البواقي المعيارية للكشف عن هذه الانحرافات.

٣-١٦-٢-١ النموذج الملائم:

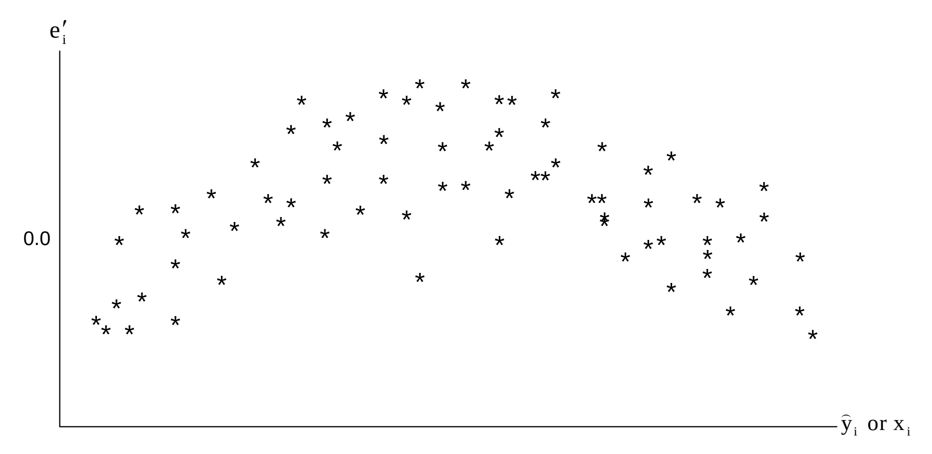
إذا كان النموذج المقدر ملائماً فإن شكل انتشار البواقي المعيارية مع القيم الموفقة أو مع أحد المتغيرات المستقلة يأخذ الشكل الموضح بالشكل رقم (٣-٣)، حيث يلاحظ أن النقط تتبعثر عشوائياً داخل حزام أفقي وسطه الصفر وتباينه ثابت ولا توجد نتوءات أو اتجاه معين تصاعدياً كان أو تنازلياً. كما يتوقع أن يقع نحو ٩٥% من قيم البواقي المعيارية ما بين -٢ و+٢ وهي فترة ثقة ٩٥%.



شكل رقم (٣-٣): شكل انتشار البواقي المعيارية مع القيم الموفقة أو أحد المتغيرات المستقلة - (حالة نموذج ملائم)

٣-١٦-٢-٢ عدم خطية دالة الانحدار:

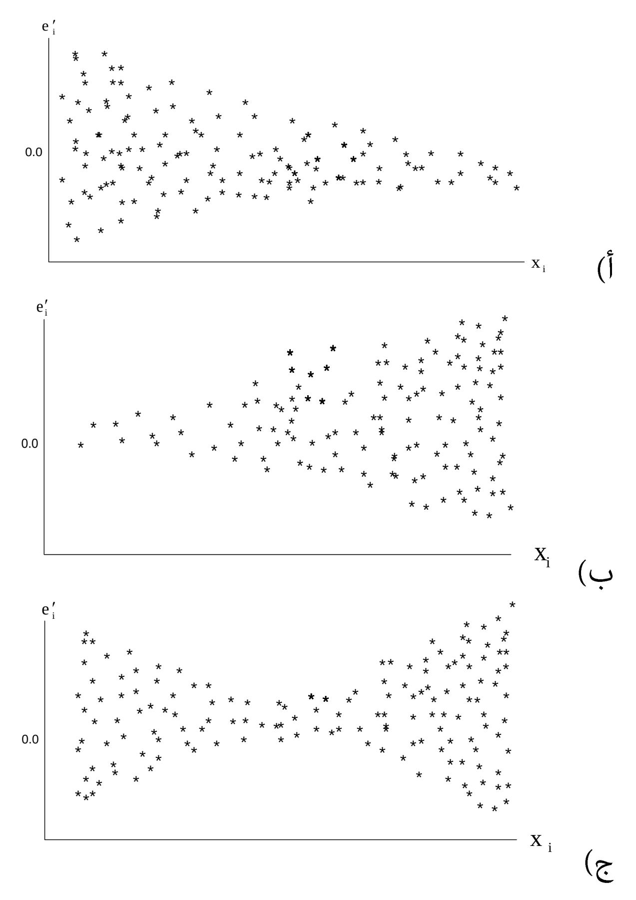
كما سبق ذكره، فإن على الباحث البدء برسم شكل انتشار المتغير التابع مع المتغيرات المستقلة لتحديد شكل العلاقة. وعلى الرغم من أنه قد يُظهر رسم شكل الانتشار أن هناك علاقة خطية تربط بين المتغير التابع والمتغير المستقل، إلا أنه بعد بناء النموذج قد يكشف لنا تحليل البواقي أن العلاقة غير خطية. ويوضح الشكل رقم (٣-٤) رسم شكل انتشار البواقي المعيارية عندما يكون النموذج الخطي غير ملائم، حيث يلاحظ أن هناك علاقة تربيعية تربط بين المتغير التابع والمتغير المستقل مما يعني ضرورة إضافة حد تربيع المتغير المستقل (x^2) للنموذج ليصبح النموذج نموذج خطي متعدد يضم (x) و(x^2).



شكل رقم (٣-٤): شكل انتشار البواقي المعيارية مع المتغير المستقل (X) أو القيم الموفقة ($\hat{y}_i$) - (حالة علاقة غير خطية)

٣-١٦-٢-٣ اختلاف التباين:

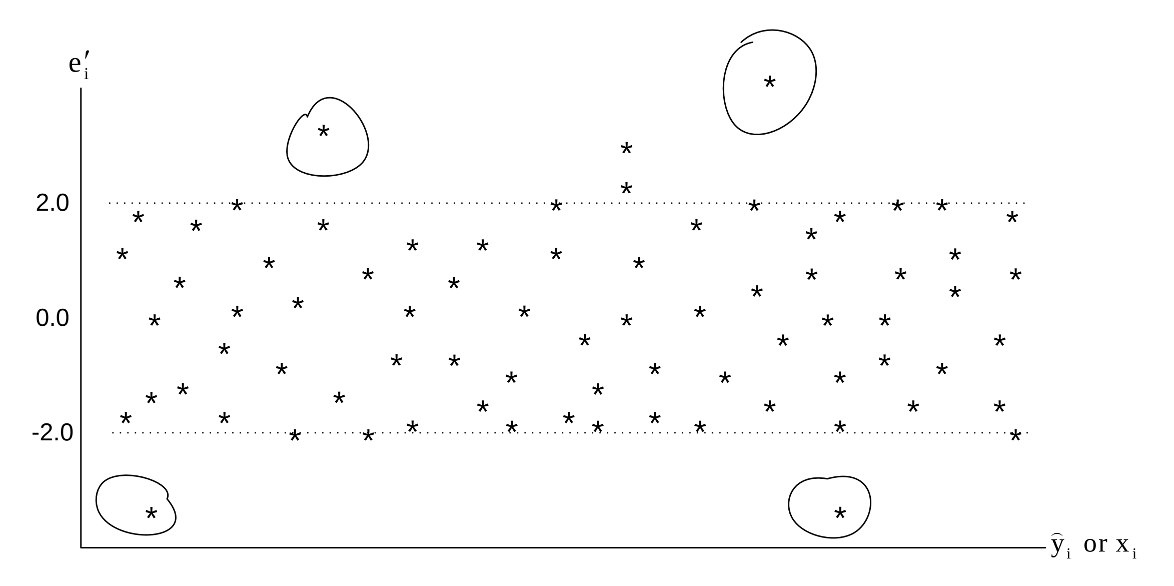
يشير اختلاف التباين إلى الحالة التي يكون فيها تباين حد الخطأ العشوائي (ε_i) غير ثابت عند كل قيم المتغير أو المتغيرات المستقلة. ويترتب على عدم ثبات تباين حدود الخطأ أن تكون مقدرات المربعات الصغرى غير كفؤة (انظر الفصل السابع). ويستخدم رسم شكل انتشار البواقي المعيارية مع أحد المتغيرات المستقلة من بين الطرق البيانية والحسابية للكشف عن عدم ثبات التباين. فإذا بدا لنا من شكل الانتشار أن تشتت قيم البواقي المعيارية يزداد بزيادة قيم أحد المتغيرات المستقلة أو عندما تزداد قيم أحد المتغيرات المستقلة يتناقص تشتت قيم البواقي المعيارية فيعني ذلك وجود اختلاف في تباين حدود الخطأ. وتوضح الأشكال رقم (٣-٥-أ)، (٣-٥-ب) و(٣-٥-ج) وجود حالات مختلفة من عدم ثبات التباين، حيث يبين الشكل رقم (٣-٥-أ) أن تباين حدود الخطأ يزداد مع زيادة قيم المتغير المستقل في حين يوضح الشكل رقم (٣-٥-ب) أن تباين حدود الخطأ يتناقص مع زيادة قيم المستقل، أما الشكل رقم (٣-٥-ج) يظهر لنا أن تباين حدود الخطأ يتناقص حتى يصل إلى أقل حد له عند القيم المتوسطة للمتغير المستقل ثم يبدأ تدريجياً في الزيادة حتى يصل إلى أعلى حد له عند القيم الكبيرة للمتغير المستقل.



شكل رقم (٣-٥): انتشار البواقي المعيارية مع أحد المتغيرات المستقلة -(حالات وجود اختلاف تباين)

٣-١٦-٢-٤ الكشف عن المشاهدات الشاذة/ الخارجة:

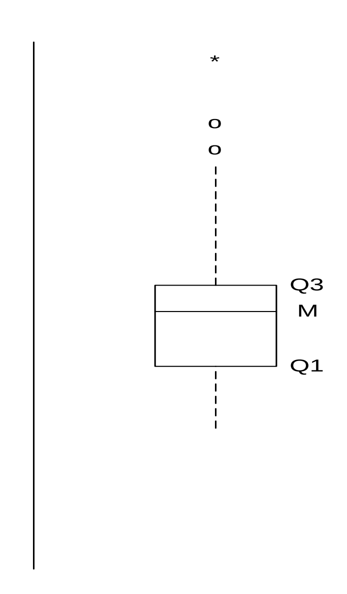
المشاهدات الشاذة أو الخارجة (Outlying Observations) هي مجموعة قليلة من المشاهدات تبعد قيمها بصورة كبيرة عن بقية قيم المشاهدات في العينة. وإن وجود بيانات شاذة في العينة قد يؤدي إلى التوصل إلى نتائج خاطئة، حيث تتأثر قيم مقدرات المربعات الصغرى بصورة كبيرة خاصة إذا كان حجم العينة (عدد المشاهدات) صغيراً. وللكشف عن البيانات الشاذة يستخدم شكل انتشار البواقي المعيارية مع القيم الموفقة ($\hat{y}_i$) أو مع المتغير المستقل. فإذا بدا من شكل الانتشار أن هناك نقطة أو عدة نقاط تبعد بصورة واضحة عن بقية النقاط، فإن هذه النقطة أو النقاط تمثل بيانات مشاهدات شاذة تستدعي دراستها. ويبين الشكل رقم (٣-٦) شكل انتشار البواقي المعيارية مع القيم الموفقة أو أحد المتغيرات المستقلة حيث يلاحظ أن هناك أربع حالات تبعد قيمها عن بقية الحالات الأخرى.



شكل رقم (٣-٦): انتشار البواقي المعيارية مع المتغير المستقل أو القيم الموفقة (حالة وجود مشاهدات شاذة)

كما يستخدم الرسم الصندوقي (Boxplot) للبواقي أيضاً للكشف عن المشاهدات الشاذة. ويتكون الرسم الصندوقي من مستطيل (صندوق) يمثل طوله الفرق (المسافة) بين قيمتي الرّبيع الثالث (Q3) والرّبيع الأول (Q1)، أو ما يعرف بالمدى الرّبيعي (Interquartile range). ويُرسم عادة خطًّا أو نقطة داخل المستطيل ليمثل الوسيط (M) يكون موقعه في المنتصف إذا كان توزيع المتغير متماثلاً (Bell-shaped) وفي مكان غير ذلك إذا كان التوزيع ملتوياً. وتقع ٥٠% من المشاهدات داخل المستطيل، أي بين الربيع الأول والربيع الثالث. وللشكل شاربان (whiskers) يتم رسمهما في شكل خط من أعلى الصندوق ومن أسفل الصندوق، يمتد طول كل منهما ١,٥ من طول المدى الرّبيعي، أي طول الخط الأعلى مساو لـ $Q3+1.5(Q3-Q1)$ والخط الأسفل مساو لـ $Q1-1.5(Q3-Q1)$. وتعتبر النقاط (المشاهدات) متطرفة (Extreme) إذا كانت قيمها أكبر من ثلاثة أمثال طول الصندوق محسوبة من الربيع الأول أو الربيع الثالث، أي أن تكون المشاهدات أكبر من $Q3+3(Q3-Q1)$ أو أن تكون المشاهدات أقل من $Q1-3(Q3-Q1)$. وتظهر

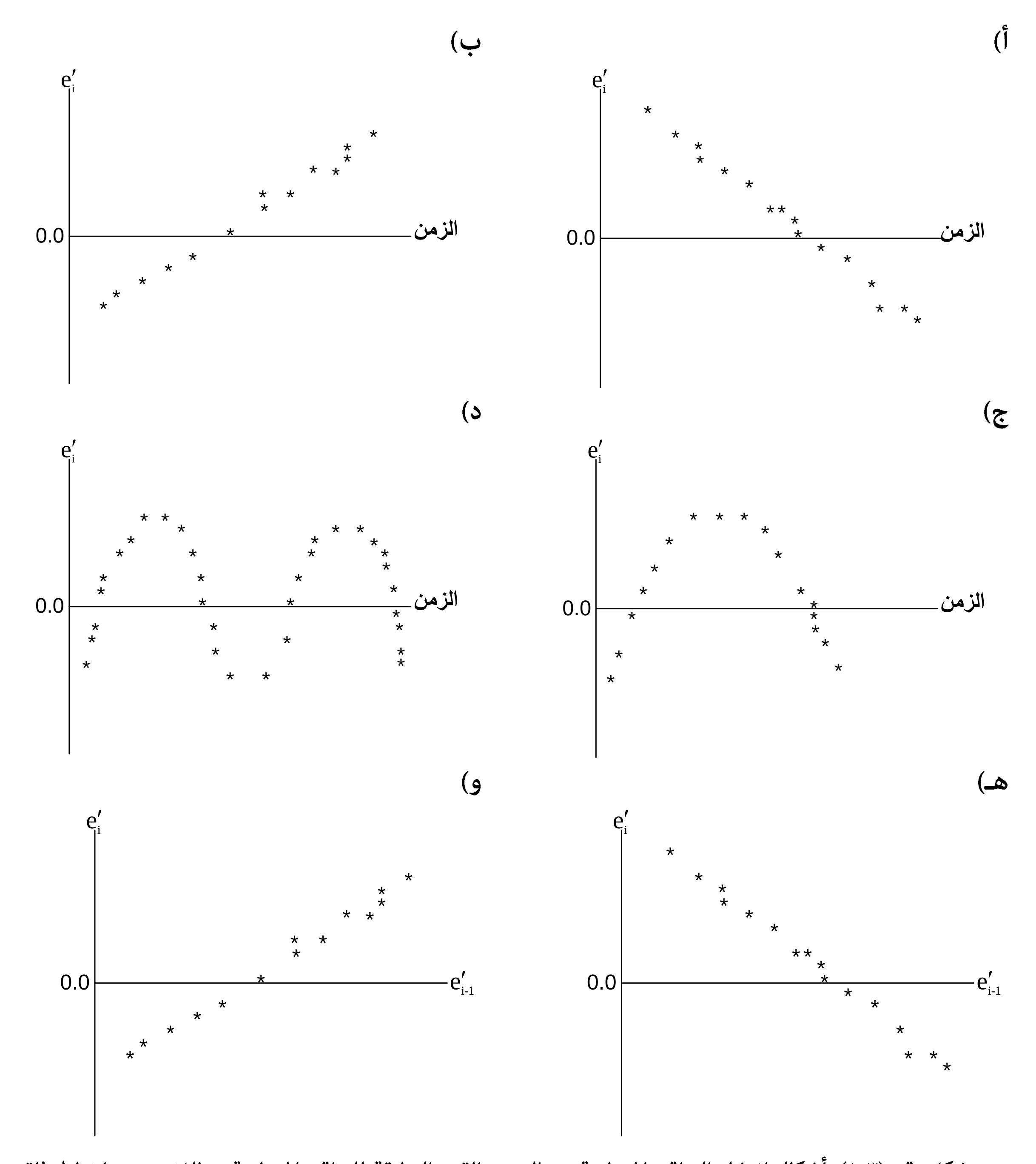
المشاهدات المتطرفة عادة في شكل نجمة (*). أما المشاهدات الشاذة (Outlier) فهي التي تبعد قيمها ما بين ١,٥ و٣ أمثال طول الصندوق محسوبة من الربيع الأول أو الربيع الثالث وتظهر عادة في شكل الحرف الإنجليزي (O) للإشارة إلى أنها شاذة (Outlier)، انظر الشكل رقم (٣-٧).



شكل رقم (٣-٧): الرسم الصندوقي للبواقي المعيارية (حالة وجود مشاهدات متطرفة ومشاهدات شاذة)

٣-١٦-٢-٥ عدم استقلالية حدود الخطأ:

يشير عدم استقلالية حدود الخطأ أو بما يعرف بالارتباط الذاتي إلى وجود ارتباط بين حدود الخطأ المتتالية. وتحدث هذه المشكلة بشكل متكرر عندما يكون المتغير التابع سلسلة زمنية. ويترتب على وجود مشكلة الارتباط الذاتي أن تكون مقدرات المربعات الصغرى غير كفؤة، وبالتالي فإن الاختبارات الإحصائية وفترات الثقة تكون خاطئة. ويستخدم رسم انتشار البواقي المعيارية مع الزمن من بين الطرق الأخرى للكشف عن وجود هذه المشكلة. فإذا كانت قيم البواقي المعيارية تزداد أو تتناقص مع الزمن دل ذلك على وجود ارتباط ذاتي بين البواقي. وتوضح الأشكال رقم (٣-٨-أ)، (٣-٨-ب)، (٣-٨-ج) و(٣-٨-د) وجود حالات مختلفة من الارتباط الذاتي بين البواقي المعيارية، حيث يتضح من الشكلين رقم (أ) و(ب) أن هناك اتجاهًا تصاعدياً وتنازلياً على التوالي في قيم البواقي المعيارية عبر الزمن، بينما يوضح الشكل رقم (ج) أن هناك ارتباطا ذاتياً بين البواقي المعيارية تأخذ صفة التقلب الدوري حيث تتغير إشارة الارتباط الذاتي من فترة زمنية إلى أخرى. أما الشكل رقم (د) فيشير إلى أن هناك علاقة خطية وتربيعية بين البواقي المعيارية. كما يستخدم رسم شكل انتشار البواقي المعيارية مع البواقي المعيارية في فترة سابقة (e'_{i-1}) للكشف عن مشكلة الارتباط الذاتي. فإذا كانت القيم الموجبة للبواقي المعيارية تتبعها قيمٌ موجبة والسالبة تتبعها قيم سالبة في معظم الحالات كما في الشكلين رقم (٣-٨-هـ) و(٣-٨-و) دل ذلك أيضاً على وجود ارتباط ذاتي بين البواقي.



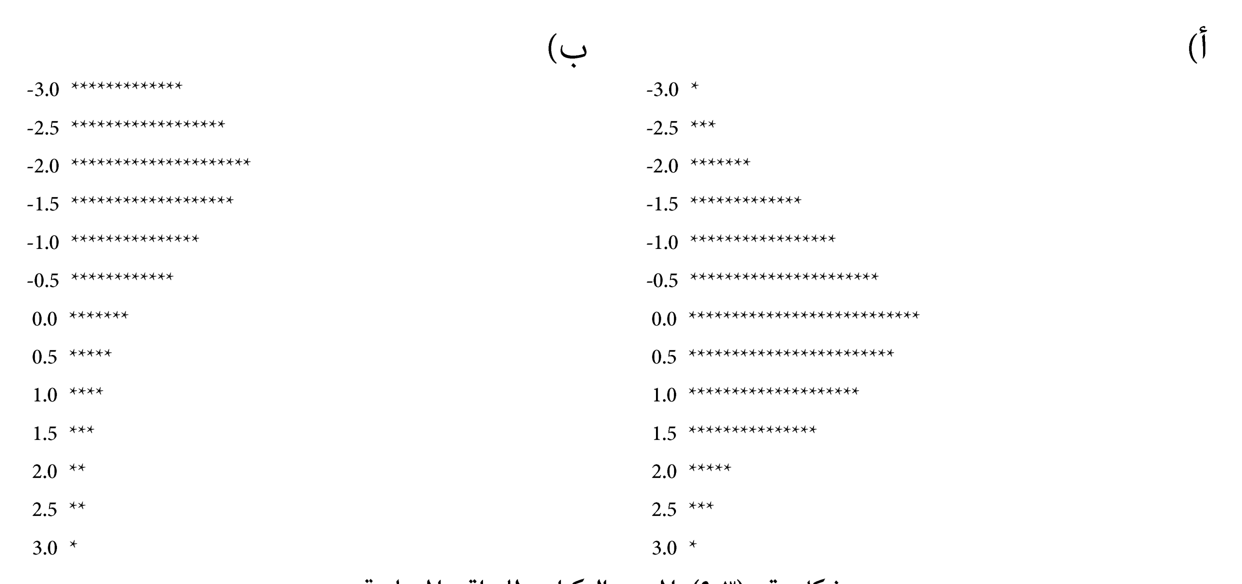
شكل رقم (٣-٨): أشكال انتشار البواقي المعيارية مع الزمن والقيم السابقة للبواقي المعيارية- حالات وجود ارتباط ذاتي

٣-١٦-٢-٦ الكشف عن تبعية البواقي للتوزيع الطبيعي:

من الاشتراطات الأساسية لتحليل الانحدار الخطي أن يكون حد الخطأ العشوائي موزعاً توزيعاً طبيعياً. وباستيفاء هذا الاشتراط يمكننا إجراء الاستدلال الإحصائي لمقدرات المربعات الصغرى. وفيما يلي نستعرض بعض الطرق البيانية المستخدمة للكشف عن اعتدالية البواقي.

المدرج التكراري للبواقي أو البواقي المعيارية:

من رسم المدرج التكراري للبواقي أو البواقي المعيارية يمكننا تحديد ما إذا كان الشكل يشبه التوزيع الطبيعي، أي ناقوسيا أم لا. ويوضح الشكل رقم (٣-٩-أ) أن البواقي المعيارية تتبع التوزيع الطبيعي، في حين يظهر الشكل رقم (٣-٩-ب) أن هناك انحرافاً عن التوزيع الطبيعي.



شكل رقم (٣-٩): المدرج التكراري للبواقي المعيارية

رسم الاحتمال الطبيعي (Normal Probability Plot):

إن رسم الاحتمال الطبيعي هو الأكثر استخداماً للكشف عن مدى تبعية البواقي للتوزيع الطبيعي. وتتلخص طريقة رسم الاحتمال الطبيعي في الخطوات التالية:

- يتم أولاً ترتيب البواقي المعيارية تصاعدياً كما يلي:

$$e'_{(1)} \; e'_{(2)} \; e'_{(3)} \; \ldots \; e'_{(n)}$$

حيث يشير الدليل السفلي بين القوسين (i) على الترتيب وn عدد المشاهدات. أي أن $e'_{(1)}$ هي أصغر البواقي قيمة و $e'_{(n)}$ هي أكبرها.

- حساب الدرجات الطبيعية (Normal Scores) على النحو التالي:

  - حساب النسب التراكمية• المناظرة للبواقي المعيارية كما يلي:

$$\left(\frac{i-0.375}{n+0.25}\right) \qquad i = 1, 2, \dots, n$$

حيث إن i هو الترتيب وn عدد المشاهدات.

- حساب الدرجات المعيارية (N_i) بإيجاد معكوس دالة التوزيع التراكمي (Inverse Cumulative distribution function of the standard normal distribution) وذلك باستخدام جدول التوزيع الطبيعي المعياري، أي:

$$N_i = \Phi_i^{-1}\left(\frac{i-0.375}{n+0.25}\right) \qquad i = 1, 2, \dots, n \tag{3-63}$$

كما يستخدم أيضاً التقريب التالي لحساب الدرجات المعيارية (Rayan and Joiner, 1976):

$$N_i = 4.91\left[\left(\frac{i-0.375}{n+0.25}\right)^{0.14} - \left\{1-\left(\frac{i-0.375}{n+0.25}\right)^{0.14}\right\}\right], i = 1, 2, \dots, n \tag{3-64}$$

- رسم شكل انتشار البواقي المعيارية (e'_i) مع الدرجات الطبيعية (N_i). فإذا بدا لنا من الشكل أن النقاط تقع تقريباً على خط مستقيم دل ذلك على أن البواقي تتبع التوزيع الطبيعي. وأما إذا أظهر الشكل خلاف ذلك فإننا نشك في تبعية البواقي للتوزيع الطبيعي.

- بالإضافة لرسم الاحتمال الطبيعي يتم حساب معامل الارتباط الخطي البسيط بين البواقي المعيارية (e'_i) والدرجات الطبيعية (N_i). فإذا كان حجم معامل الارتباط كبيراً دل ذلك على تبعية البواقي أو البواقي المعيارية للتوزيع الطبيعي. ويرفض فرض تبعية البواقي للتوزيع الطبيعي عند مستوى دلالة (α) إذا كانت قيمة معامل الارتباط أقل من القيم المناظرة في ملحق الجداول الإحصائية.

- **اختبار جارك –بيرا (Jarque-Bera Test)**: يعد اختبار جارك وبيرا (Jarque & Bera, 1980, 1987) من الاختبارات الواسعة الاستخدام لاختبار تبعية البواقي للتوزيع الطبيعي. والفرض المطلوب اختباره كالتالي:

• يستخدم بعض الكتاب نسبًا تراكمية أخرى مثل $\frac{(i-0.33)}{(n+0.33)}$ (**انظر** Toit,Steyn & Stumpf, 1986, pp 40-45)

فرض العدم: البواقي تتبع التوزيع الطبيعي مقابل **الفرض البديل**: البواقي لا تتبع التوزيع الطبيعي

ويأخذ إحصاء الاختبار الصيغة التالية:

$$JB = \frac{n}{6}\left(SK^2 + \frac{(KR\text{-}3)^2}{4}\right) \sim \chi^2_2 \qquad (3\text{-}65)$$

حيث إن:

n عدد المشاهدات.

SK معامل الالتواء وصيغته $SK = \frac{\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2\right)^{3/2}}$.

KR معامل التفرطح وصيغته $KR = \frac{\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2\right)^2}$

ويتبع إحصاء الاختبار توزيع مربع كاي بدرجتي حرية، حيث يتم رفض فرض العدم إذا كانت قيمة الإحصاء أكبر من القيمة الحرجة لتوزيع مربع كاي (Chi-square χ^2) بدرجتي حرية ومستوى دلالة محدد (α=0.05 مثلاً).

٣-١٦-٣ مثال:

بالنسبة لنموذج انحدار وزن الطفل على عمر وطول الطفل، احسب البواقي والبواقي المعيارية ومن ثم افحص النموذج للتأكد من مدى استيفائه للاشتراطات الأساسية باستخدام تحليل البواقي؟

الحل:

- البواقي

كما سبق ذكره، البواقي هي الفروق بين القيم الفعلية والقيم المقدرة، أي:

$$e_i = y_i - \hat{y}_i$$
$$e_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2\right)$$

وبما أن نموذج الإنحدار الموفق هو $\hat{y}$ = -2.1819 + 1.2008x_1+0.12457x_2 ، فإن:

$$e_i = y_i + 2.1819 - 1.2008x_1 - 0.12457x_2$$

فمثلاً بالنسبة للمشاهدات الثلاث الأولى يتم حساب البواقي على النحو التالي (انظر الجدول رقم (٣-٤)):

$$e_1 = 11.50 + 2.1818689 - 1.20080469 \times 3 - 0.1245725 \times 84 = -0.38464$$
$$e_2 = 16.00 + 2.1818689 - 1.20080469 \times 5 - 0.1245725 \times 95 = 0.34346$$
$$e_3 = 6.50 + 2.1818689 - 1.20080469 \times 0.5 - 0.1245725 \times 65 = -0.01575$$

البواقي المعيارية:

لحساب البواقي المعيارية تتبع الخطوات التالية:

- حساب مصفوفة القبعة ($\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T$) كما يلي:

$$\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T = \begin{pmatrix} 1 & 3.00 & 84.0 \\ 1 & 5.00 & 95.0 \\ 1 & 0.50 & 65.0 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & 0.08 & 40.0 \end{pmatrix} \begin{pmatrix} 0.789947 & 0.144238 & -0.014059 \\ 0.144238 & 0.036936 & -0.0030227 \\ -0.014509 & -0.003023 & 0.0002828 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & . & . & . & 1 \\ 3.0 & 5.0 & 0.5 & . & . & . & 0.08 \\ 84.0 & 95.0 & 65.0 & . & . & . & 40.0 \end{pmatrix}$$

$$\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T = \begin{pmatrix} 0.0220803 & 0.0262932 & 0.0158908 & . & . & . & 0.011015 \\ 0.0262932 & 0.0794659 & -0.0256881 & . & . & . & 0.025747 \\ 0.0158908 & -0.0256881 & 0.0554751 & . & . & . & 0.010661 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0.0110153 & 0.0257466 & 0.0106611 & . & . & . & 0.085621 \end{pmatrix}$$

ومن مصفوفة القبعة أعلاه يمكن إيجاد قيم h_{ii} وقيم $1\text{-}h_{ii}$ للمشاهدات الثلاث الأُوَل كما يلي:

$h_{11} = 0.02208$ $\quad$ $1\text{-}h_{11} = 0.97792$

$h_{22} = 0.0794659$ $\quad$ $1\text{-}h_{22} = 0.9205341$

$h_{33} = 0.055475$ $\quad$ $1\text{-}h_{33} = 0.944525$

ومن جدول تحليل التباين نجد التباين (S^2) يساوى (1.2063)، وبالتعويض المباشر في معادلة البواقي المعيارية (3.62) يمكننا مثلاً حساب البواقي المعيارية للمشاهدات الثلاث الأُوَل كما يلي:

حيث إن:

$$e_i' = \frac{e_i}{\left\{\left(1-h_{ii}\right)S^2\right\}^{1/2}}$$

فإن:

$$e_1' = \frac{-0.384636}{\sqrt{[(1\text{-}0.02208)\times 1.2062958)]}} = -0.354137$$

$$e_2' = \frac{0.34346}{\sqrt{[(1\text{-}0.079466)\times 1.2062958)]}} = 0.3259130$$

وتشير القيم المطلقة الكبيرة للبـواقي المعياريـة إلى أن المشـاهدة متطرفـة أو شـاذة. فمـثلاً قـيم البـواقي المعياريـة المقابلة للحالات (٢٠)، (٤٥)، و(٤٦) تعتبر قيم كبيرة يسترعي دراستها (انظر الفصل الرابع).

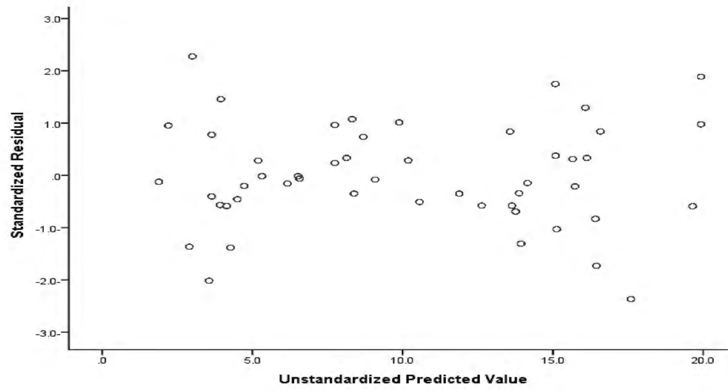
جدول رقم (٣-٤): البواقي والبواقي المعيارية لنموذج وزن الطفل

المشاهدة	البواقي e_i	h_{ii}	$1\text{-}h_{ii}$	البواقي المعيارية e_i'
1	-0.384636	0.022080	0.977920	-0.354137
2	0.343457	0.079466	0.920534	0.325930
3	-0.015747	0.055475	0.944525	-0.014752
4	1.921399	0.042587	0.957413	1.787890
5	0.364723	0.030444	0.969556	0.337248
6	1.060988	0.046495	0.953505	0.989287
7	2.073347	0.087742	0.912258	1.976452
8	-0.759272	0.039708	0.960292	-0.705454
9	-1.434901	0.031478	0.968522	-1.327517
10	-0.910026	0.106077	0.893923	-0.876349
11	0.809416	0.068384	0.931616	0.763531
12	0.415351	0.041720	0.958280	0.386315
13	1.183134	0.050258	0.949742	1.105361
14	0.922405	0.126020	0.873980	0.898346
15	-0.649000	0.093196	0.906804	-0.620527
16	-0.086849	0.044507	0.955493	-0.080896
17	-1.129657	0.037194	0.962806	-1.048216
18	0.314458	0.035674	0.964326	0.291557
19	-0.170537	0.051839	0.948161	-0.159460
20	-2.598735	0.078368	0.921632	-2.464657
21	-0.634700	0.035957	0.964043	-0.588563
22	1.073347	0.087742	0.912258	1.023186
23	-0.632072	0.034371	0.965629	-0.585645
24	0.920393	0.074578	0.925422	0.871117
25	0.309543	0.039202	0.960798	0.287526
26	-0.015030	0.040142	0.959858	-0.013968
27	-0.066803	0.035613	0.964387	-0.061936

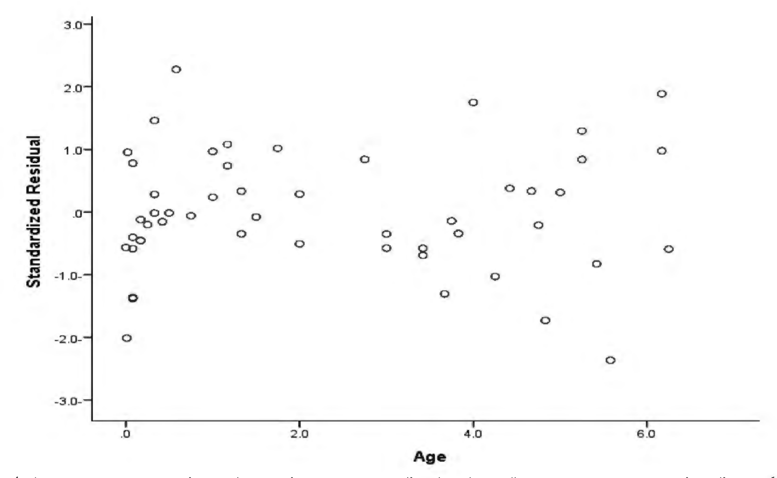
المشاهدة	البواقي e_i	h_{ii}	$1-h_{ii}$	البواقي المعيارية e'_i
28	-0.377884	0.030187	0.969813	-0.349372
29	-0.220676	0.040615	0.959385	-0.205131
30	-0.230060	0.051419	0.948581	-0.215068
31	0.367714	0.045121	0.954879	0.342618
32	1.114660	0.049835	0.950165	1.041156
33	1.418683	0.095908	0.904092	1.358477
34	-1.898987	0.048511	0.951489	-1.772528
35	-0.559259	0.050623	0.949377	-0.522597
36	-0.500039	0.042282	0.957718	-0.465220
37	-0.642821	0.045099	0.954901	-0.598941
38	0.260988	0.046495	0.953505	0.243351
39	-0.384422	0.036634	0.963366	-0.356604
40	-0.155538	0.032776	0.967224	-0.143994
41	-0.134016	0.180373	0.819627	-0.134779
42	-0.444531	0.054521	0.945479	-0.416245
43	1.605268	0.060904	0.939096	1.508223
44	-1.517394	0.044157	0.955843	-1.413117
45	-2.210475	0.053561	0.946439	-2.068771
46	2.500792	0.165276	0.834724	2.492178
47	0.855469	0.054521	0.945479	0.801035
48	1.047815	0.122255	0.877745	1.018295
49	-0.622184	0.046989	0.953011	-0.580288
50	-1.497096	0.085621	0.914379	-1.425475

فحص النموذج:

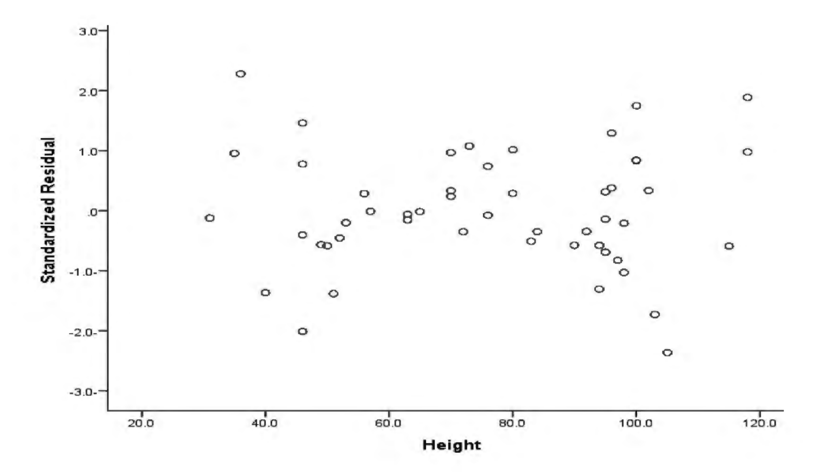
تبين الأشكال رقم (٣-١٠)، (٣-١١)، و(٣-١٢) رسم انتشار البواقي المعيارية مع القيم الموفقة ($\hat{y}_i$)، متغير العمر، والطول على التوالي. ويلاحظ من هذه الأشكال أن النقط تتبعثر عشوائياً ولا توجد نتوءات باستثناء نقطتين تبعد قليلاً عن بقية النقط، مما يدل على أن اشتراطات المربعات الصغرى مستوفية. وللكشف عن اعتدالية البواقي، يوضح رسم المدرج التكراري أن البواقي المعيارية لها تقريباً توزيع طبيعي؛ ذلك لأن النقاط لها شكل شبيه بشكل الجرس (شكل رقم (٣-١٣)). وللتحقق أكثر عن مدى تبعية البواقي للتوزيع الطبيعي تم رسم الاحتمال الطبيعي. حيث تم أولاً ترتيب البواقي المعيارية ترتيباً تصاعدياً ومن ثم تم حساب النسب التراكمية والدرجات المعيارية المناظرة لها (انظر الجدول رقم (٣-٥)). وبتوقيع الدرجات المعيارية مع البواقي المعيارية نلاحظ أن أزواج القيم (N_i، e'_i) تقع تقريباً على خط مستقيم وبالتالي يمكننا القول إن بواقي نموذج وزن الطفل تتبع التوزيع الطبيعي (شكل رقم (٣-١٤)). ويدعم هذه النتيجة معامل ارتباط البواقي المعيارية مع الدرجات الطبيعية الذي بلغ (٠,٩٩٣)، أكبر بكثير من القيم الحرجة (٠,٩٧٦) عند مستوى معنوية (٠,٠٥)، مما يعني البواقي المعيارية تتبع التوزيع الطبيعي.



شكل رقم (٣-١٠): رسم البواقي المعيارية مع القيم الموفقة لنموذج انحدار وزن الطفل



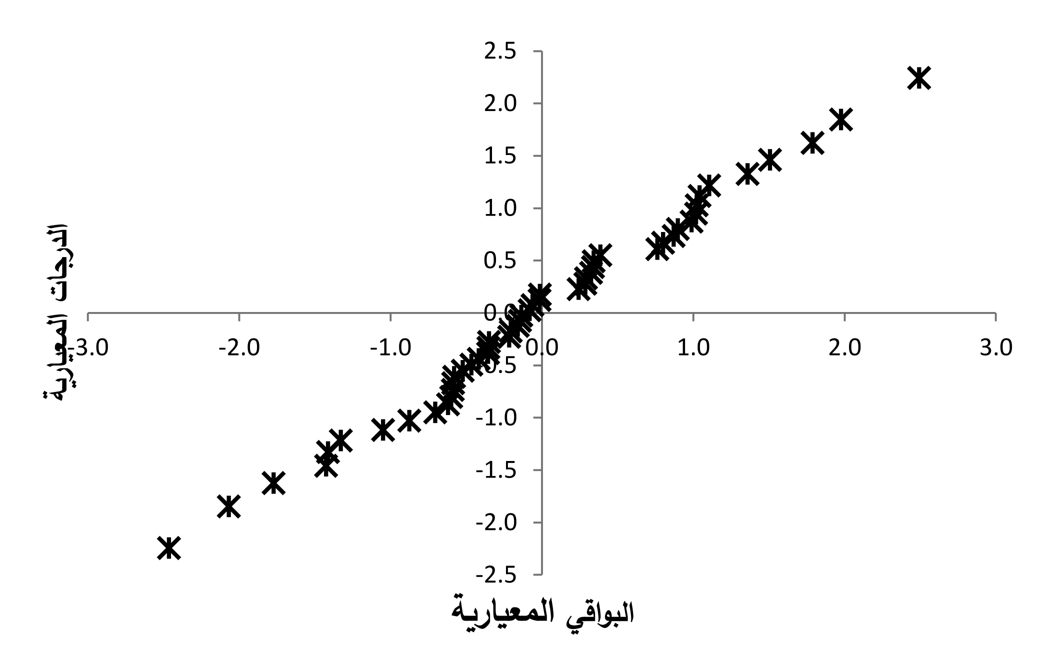
شكل رقم (٣-١١): رسم البواقي المعيارية مع متغير العمر لنموذج انحدار وزن الطفل



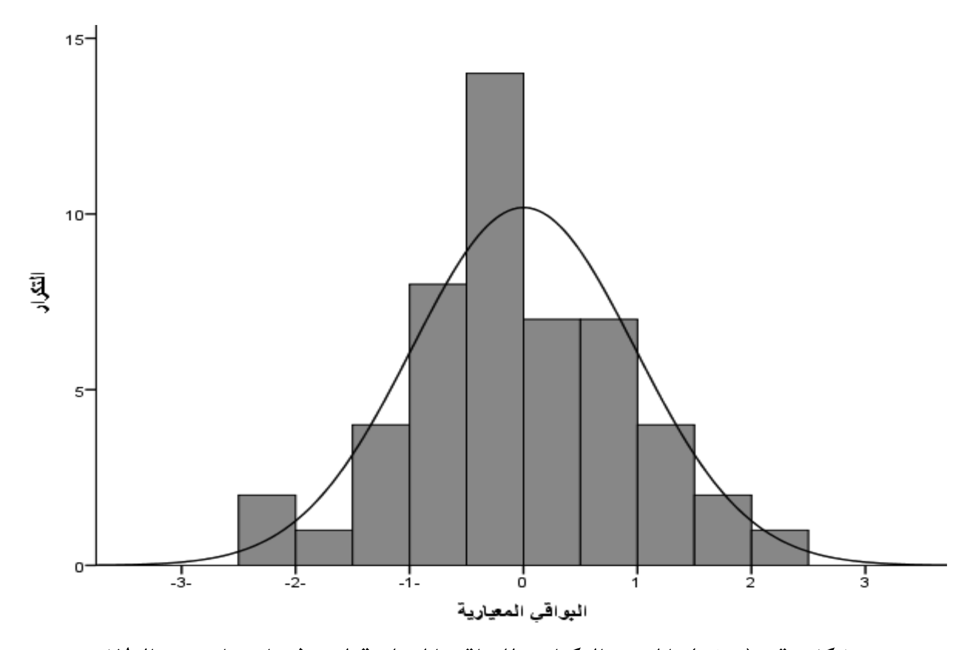
شكل رقم (٣-١٢): رسم البواقي المعيارية مع متغير الطول لنموذج انحدار وزن الطفل

جدول رقم (٣-٥): البواقي المعيارية، النسب التراكمية، الدرجات المعيارية لنموذج وزن الطفل

البواقي المعيارية	الترتيب	النسب التراكمية $\frac{(i-0.375)}{50+0.25}$	الدرجة المعيارية
-2.464657	1	0.01244	-2.24333
-2.068771	2	0.03234	-1.84749
-1.772528	3	0.05224	-1.62352
-1.425475	4	0.07214	-1.46004
-1.413117	5	0.09204	-1.32830
-1.327517	6	0.11194	-1.21627
-1.048216	7	0.13184	-1.11773
-0.876349	8	0.15174	-1.02899
-0.705454	9	0.17164	-0.94770
-0.620527	10	0.19154	-0.87223
-0.598941	11	0.21144	-0.80143
-0.588563	12	0.23134	-0.73443
-0.585645	13	0.25124	-0.67058
-0.580288	14	0.27114	-0.60936
-0.522597	15	0.29104	-0.55034
-0.465220	16	0.31095	-0.49317
-0.416245	17	0.33085	-0.43758
-0.356604	18	0.35075	-0.38331
-0.354137	19	0.37065	-0.33014
-0.349372	20	0.39055	-0.27789
-0.215068	21	0.41045	-0.22639
-0.205131	22	0.43035	-0.17549
-0.159460	23	0.45025	-0.12503
-0.143994	24	0.47015	-0.07489
-0.134779	25	0.49005	-0.02494
-0.080896	26	0.50995	0.02494
-0.061936	27	0.52985	0.07489
-0.014752	28	0.54975	0.12503
-0.013968	29	0.56965	0.17549
0.243351	30	0.58955	0.22639
0.287526	31	0.60945	0.27789
0.291557	32	0.62935	0.33014
0.325930	33	0.64925	0.38331
0.337248	34	0.66915	0.43758
0.342618	35	0.68905	0.49317
0.386315	36	0.70896	0.55034
0.763531	37	0.72886	0.60936
0.801035	38	0.74876	0.67058
0.871117	39	0.76866	0.73443
0.898346	40	0.78856	0.80143
0.989287	41	0.80846	0.87223
1.018295	42	0.82836	0.94770
1.023186	43	0.84826	1.02899
1.041156	44	0.86816	1.11773
1.105361	45	0.88806	1.21627
1.358477	46	0.90796	1.32830
1.508223	47	0.92786	1.46004
1.787890	48	0.94776	1.62352
1.976452	49	0.96766	1.84749
2.492178	50	0.98756	2.24333



شكل رقم (٣-١٣): رسم الاحتمال الطبيعي للبواقي المعيارية لنموذج انحدار وزن الطفل



شكل رقم (٣-١٤): المدرج التكراري للبواقي المعيارية لنموذج انحدار وزن الطفل

اختبار جارك -بيرا:

بالإضافة للاختبارات السابقة، يمكن إجراء اختبار جارك-بيرا للتأكد من تبعية البواقي للتوزيع الطبيعي. ولحساب الاختبار تم أولاً حساب معاملي الالتواء والتفرطح كما يلي (انظر الجدول ٣-٦):

معامل الالتواء:
$$SK=\frac{\frac{1}{n}\sum_{i=1}^{n}(e_i-\bar{e})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(e_i-\bar{e})^2\right)^{3/2}}=\frac{\frac{1}{50}\times-0.9996}{\left(\frac{1}{50}\times49.3346\right)^{3/2}}=-0.02061$$

معامل التفرطح:
$$KR=\frac{\frac{1}{n}\sum_{i=1}^{n}(e_i-\bar{e})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(e_i-\bar{e})^2\right)^{2}}=\frac{\frac{1}{50}\times143.7971}{\left(\frac{1}{50}\times49.3346\right)^{2}}=2.9945$$

ومن ثم يتم حساب إحصاء جارك-بيرا كما يلي:

$$JB=\frac{n}{6}\left(SK^2+\frac{(KR\text{-}3)^2}{4}\right)=\frac{50}{6}\left(-0.02061^2+\frac{(2.9945\text{-}3)^2}{4}\right)=0.0036$$

وحيث إن قيمة إحصاء جارك - بيرا أقل بكثير من القيمة الحرجة لتوزيع مربع كاي عند درجتي حرية ومستوى معنوية* ٥% $\left(\chi^2_{\alpha=0.05}=5.99\right)$، فإنه لا يوجد دليل كاف لرفض فرض العدم ومن ثم فإن البواقي أو البواقي المعيارية تتبع التوزيع الطبيعي.

* كما يمكن حساب قيمة الاحتمال باستخدام برنامج إكسل كما يلي: =chidist(0.0035;2) = 0.998

جدول رقم (٣-٦): الحسابات اللازمة لحساب معاملي الالتواء والتفرطح

المشاهدة	الباقي المعياري e_i	$(e_i - \bar{e})^2$	$(e_i - \bar{e})^3$	$(e_i - \bar{e})^4$
1	-0.3576	0.1279	-0.0457	0.0163
2	0.3193	0.4365	0.0326	0.0104
3	-0.0146	0.0002	0.0000	0.0000
4	1.7862	3.1907	5.6993	10.1803
5	0.3391	0.1150	0.0390	0.0132
6	0.9864	0.9729	0.9596	0.9465
7	1.9275	3.7153	7.1611	13.8031
8	-0.7059	0.4982	-0.3517	0.2482
9	-1.3340	1.7795	-2.3737	3.1665
10	-0.8460	0.7157	-0.6055	0.5123
11	0.7525	0.5662	0.4261	0.3206
12	0.3861	0.1491	0.0576	0.0222
13	1.0999	1.2098	1.3307	1.4636
14	0.8575	0.7353	0.6306	0.5407
15	-0.6033	0.3640	-0.2196	0.1325
16	-0.0807	0.0065	-0.0005	0.0000
17	-1.0502	1.1029	-1.1583	1.2164
18	0.2923	0.0855	0.0250	0.0073
19	-0.1585	0.0251	-0.0040	0.0006
20	-2.4159	5.8367	-14.1011	34.0672
21	-0.5901	0.3482	-0.2054	0.1212
22	0.9978	0.9957	0.9935	0.9914
23	-0.5876	0.3453	-0.2029	0.1192
24	0.8556	0.7321	0.6264	0.5360
25	0.2878	0.0828	0.0238	0.0069
26	-0.0140	0.0002	0.0000	0.0000
27	-0.0621	0.0039	-0.0002	0.0000
28	-0.3513	0.1234	-0.0434	0.0152
29	-0.2052	0.0421	-0.0086	0.0018
30	-0.2139	0.0457	-0.0098	0.0021
31	0.3418	0.1169	0.0399	0.0137
32	1.0362	1.0738	1.1127	1.1531
33	1.3189	1.7395	2.2942	3.0257
34	-1.7654	3.1167	-5.5022	9.7135
35	-0.5199	0.2703	-0.1405	0.0731
36	-0.4649	0.2161	-0.1005	0.0467
37	-0.5976	0.3571	-0.2134	0.1275
38	0.2426	0.0589	0.0143	0.0035
39	-0.3574	0.1277	-0.0456	0.0163
40	-0.1446	0.0209	-0.0030	0.0004
41	-0.1246	0.0155	-0.0019	0.0002
42	-0.4133	0.1708	-0.0706	0.0292
43	1.4923	2.2271	3.3236	4.9600
44	-1.4107	1.9899	-2.8071	3.9599
45	-2.0550	4.2229	-8.6781	17.8333
46	2.3249	5.4050	12.5661	29.2145
47	0.7953	0.6325	0.5030	0.4000
48	0.9741	0.9489	0.9243	0.9004
49	-0.5784	0.3346	-0.1935	0.1119
50	-1.3918	1.9371	-2.6960	3.7522
المجموع	0.0000	49.3346	-0.9996	143.7971

٣-١٦-٤ رسوم الانحدار الجزئية (Partial Regression Plots):

تُعَدُّ رسوم الانحدار الجزئية من الأدوات المهمة المكملة لتحليل البـواقي والتـي تسـتخدم في الكشـف عـن طبيعـة العلاقـة الدالية بين المتغير المستقل والمتغير التابع والكشف عن المشاهدات الشاذة واختلاف التباين. والرسـم البيـاني الجـزئي لأي متغـير مستقل، X_k مثلاً، عبارة عن رسم شكل انتشار بواقي نموذج انحدار المتغير X_k على بقية المتغيرات المستقلة مـع بـواقي نمـوذج انحدار المتغير التابع على المتغيرات المستقلة باستثناء المتغير X_k ويتطلب إجراء الرسوم البيانية الجزئية أن يكون عدد المتغـيرات المستقلة اثنين فأكثر. وفيما يلي الخطوات التي توضح كيفية إجراء الرسم البياني الجزئي:

- يأخذ نموذج الانحدار الخطي المتعدد الذي يضم عدد (p) متغير مستقل الصيغة التالية:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i$$

ولإجراء الرسم البياني الجزئي للمتغير X_1 مثلاً، يتم أولاً إجراء نموذج انحدار Y عـلى كـل المتغـيرات المسـتقلة مـا عـدا المتغير X_1 ويتم الحصول على القيم الموفقة والبواقي على النحو التالي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_p x_{pi}$$

$$e_i(y) = y_i - \hat{y}_i \qquad (3\text{-}66)$$

- في الخطوة الثانية يتم إجراء نموذج انحدار المتغير X_1 على بقية المتغيرات المسـتقلة ويـتم الحصـول عـلى القـيم الموفقة والبواقي كما يلي:

$$\hat{x}_{1i} = \ddot{\beta}_0 + \ddot{\beta}_2 x_{2i} + \ldots + \ddot{\beta}_p x_{pi}$$

$$e_i(X_1) = x_{1i} - \hat{x}_{1i} \qquad (3\text{-}67)$$

- في الخطوة الثالثة يتم رسم شكل انتشار بواقي نموذج انحدار Y على جميع المتغيرات المسـتقلة باسـتثناء المتغـير X_1 حسب المعادلة (3.66) مع بواقي نموذج انحدار المتغير X_1 على بقية المتغيرات المستقلة حسب المعادلـة (3.67). والشكل الناتج هو رسم الانحدار الجزئي للمتغير المستقل X_1. وباتباع نفس الطريقة يمكن إجـراء رسـم الانحـدار الجزئي لبقية المتغيرات المستقلة $(X_2, X_3, \ldots, X_p)$.

وللبواقي $e_i(Y)$ و $e_i(X_k)$ الخصائص التالية:

- أن ميل نموذج انحدار $e_i(Y)$ على $e_i(X_k)$ هو نفس معامل المتغير X_k المقدر في نموذج الانحدار الخطي المتعدد، أي أن:

$$e_i(y) = \hat{\beta}_k e_i(x_k) + e_i$$

- أن بواقي نموذج انحدار $e_i(Y)$ على $e_i(X_k)$ هي نفس بواقي نموذج الانحدار المقدر الكامل، أي أن:

$$e_i(y) - \hat{\beta}_k e_i(x_k) = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_p x_{pi})$$

مثال:

بالنسبة لنموذج انحدار وزن الطفل على العمر والطول، ارسم رسم الانحدار الجزئي لمتغيري العمر والطول؟

الحل:

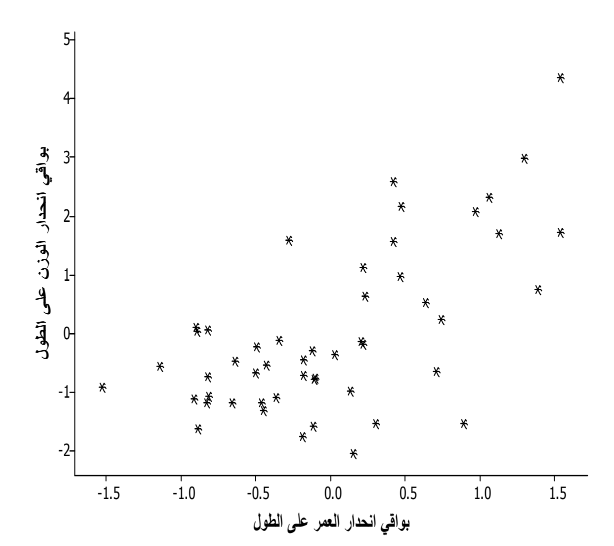
رسم الانحدار الجزئي لمتغير العمر:

- تم إجراء نموذج انحدار الوزن على الطول وتم الحصول على بواقي النموذج المقدر (جدول رقم (٣-٧)).
- تم إجراء نموذج انحدار العمر على الطول وتم الحصول على بواقي النموذج المقدر (جدول رقم (٣-٧)).

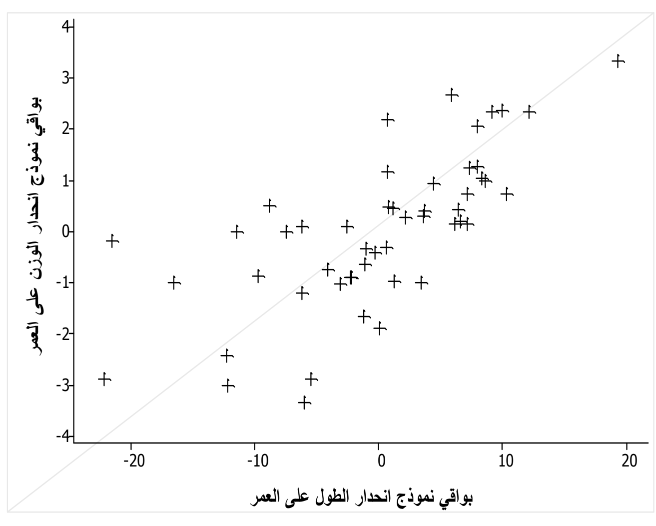
رسم الانحدار الجزئي لمتغير الطول:

- تم إجراء نموذج انحدار الوزن على العمر وتم الحصول على بواقي النموذج المقدر (جدول رقم (٣-٧)).
- تم إجراء نموذج انحدار الطول على العمر وتم الحصول على بواقي النموذج المقدر (جدول رقم (٣-٧)).

ويلاحظ أنه يوجد بعض المشاهدات قيمها تبعد قليلاً عن بقية قيم المشاهدات مما قد يشير إلى أنها شاذة. ويوضح الشكلان (٣-١٥) و(٣-١٦) أن هناك علاقة خطية تربط بين كل من المتغيرين - العمر والطول - والمتغير التابع. وعموماً يمكن القول إن أيّا من المتغيرين يسهم في تفسير التغير في وزن الطفل بوجود الآخر.



شكل رقم (٣-١٥): رسم الانحدار الجزئي لمتغير العمر



شكل رقم (٣-١٦): رسم الانحدار الجزئي لمتغير الطول

جدول رقم (٣-٧): بواقي نماذج الوزن على الطول، العمر على الطول، الوزن على العمر والطول على العمر

م	بواقي نموذج الوزن على الطول	بواقي نموذج العمر على الطول	بواقي نموذج الوزن على العمر	بواقي نموذج الطول على العمر
1	-0.3476	0.03084	-0.30715	0.622
2	1.70113	1.13063	-0.87196	-9.7567
3	-1.11359	-0.91426	1.02386	8.3454
4	1.58692	-0.27855	2.66044	5.9326
5	-0.22781	-0.49344	0.92196	4.4732
6	0.07219	-0.82344	2.05766	8.0007
7	2.57575	0.41839	2.16512	0.7367
8	-1.29887	-0.44937	0.12924	7.1325
9	-1.57603	-0.11753	-1.00387	3.4601
10	0.75544	1.38696	-2.43557	-12.2463
11	-0.56486	-1.14447	2.32715	12.1835
12	0.97829	0.4688	0.09683	-2.5569
13	0.10366	-0.89895	2.32715	9.1835
14	-0.91308	-1.52855	3.32595	19.2943
15	0.24428	0.7439	-1.03747	-3.1184
16	-1.06486	-0.81447	0.99145	8.656
17	-0.9674	0.13512	-0.97266	1.2603
18	-0.45623	-0.64181	1.22525	7.3113
19	-1.16791	-0.83059	0.72645	7.2005
20	-1.5273	0.89227	-3.34076	-5.9566
21	-1.07603	-0.36753	0.12924	6.1325
22	1.57575	0.41839	1.16512	0.7367
23	-1.18466	-0.46018	0.19285	6.622
24	2.08692	0.97145	-0.00506	-7.4291
25	-0.10801	-0.34773	0.45437	1.1626
26	-0.53085	-0.42957	0.25437	2.1626
27	-0.66791	-0.50059	0.39076	3.673
28	-0.13034	0.20615	-0.40905	-0.2502
29	-0.43948	-0.18222	-0.34304	-0.9823
30	0.5326	0.63512	-0.73886	-4.0844
31	0.64123	0.22778	0.46373	0.7708
32	0.04377	-0.89181	2.35835	9.9837
33	2.97829	1.2988	-0.00506	-11.4291
34	-1.53161	0.30594	-1.89145	0.0605
35	-1.62476	-0.88732	0.72525	10.3113
36	-0.71664	-0.18038	-0.64045	-1.1271
37	-0.77096	-0.10671	-0.91253	-2.1651
38	-0.72781	-0.82344	1.25766	8.0007
39	-1.17349	-0.65712	0.42196	6.4732
40	-0.29887	-0.11937	0.29354	3.605
41	1.71305	1.53819	-2.89045	-22.1271
42	-0.17959	0.22064	-1.21253	-6.1651
43	2.17041	0.47064	0.50437	-8.8374
44	-1.7438	-0.18854	-1.66253	-1.1651
45	-2.02959	0.15064	-2.88526	-5.4168
46	4.34884	1.53901	-0.17873	-21.5098
47	1.12041	0.22064	0.08747	-6.1651
48	2.32168	1.06085	-1.01059	-16.5237
49	-0.74811	-0.10487	-0.90994	-2.3099
50	-0.64253	0.71166	-3.01253	-12.1651

٣-١٧ نماذج الانحدار متعددة الحدود (Polynomial Regression Models):

٣-١٧-١ مقدمة:

يعرف نموذج الانحدار متعدد الحدود بأنه دالة يظهر فيها متغير مستقل أساسي عدداً من المرات مرفوعاً في كل مرة إلى درجة أعلى. ويأخذ النموذج من الدرجة (p) الصيغة التالية:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \varepsilon_i \qquad (3.68)$$

ويُعَدُّ النموذج متعدد الحدود حالة خاصة من نموذج الانحدار الخطي المتعدد. وبما أنه يوجد متغير أساسي واحد فإنه يمكن تمثيل أي نموذج متعدد الحدود برسم شكل انتشار بين المتغير التابع والمتغير المستقل بهدف تحديد منحنى يصف العلاقة بينهما.

وفي هذا الجزء من الفصل ستقتصر دراستنا على نموذجي الانحدار من الدرجة الثانية والدرجة الثالثة؛ ذلك لأن الدرجات الأعلى نادراً ما تستخدم في الواقع العملي فضلاً عن أن مشكلة الارتباط الخطي* المتعدد تتفاقم بزيادة عدد الحدود. وبما أن هذه النماذج كما أسلفنا هي حالات خاصة من نموذج الانحدار الخطي المتعدد فإنه يتم استخدام طريقة المربعات الصغرى الاعتيادية لتقدير معالم النموذج، وكذلك يتم حساب الإحصاءات الأخرى كمعامل التحديد، الخطأ المعياري وغيرها بنفس الصيغ المستخدمة لنموذج الانحدار الخطي المتعدد.

٣-١٧-٢ نموذج الانحدار من الدرجة الثانية (Quadratic Regression Model):

أ- النموذج:

يُعَدُّ نموذج الانحدار من الدرجة الثانية أو ما يعرف أيضاً بنموذج الانحدار التربيعي من أبسط أنواع الانحدار متعدد الحدود حيث يضم النموذج المتغير (X) والمتغير (X^2). ويأخذ النموذج الصيغة التالية:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \qquad (3\text{-}69)$$

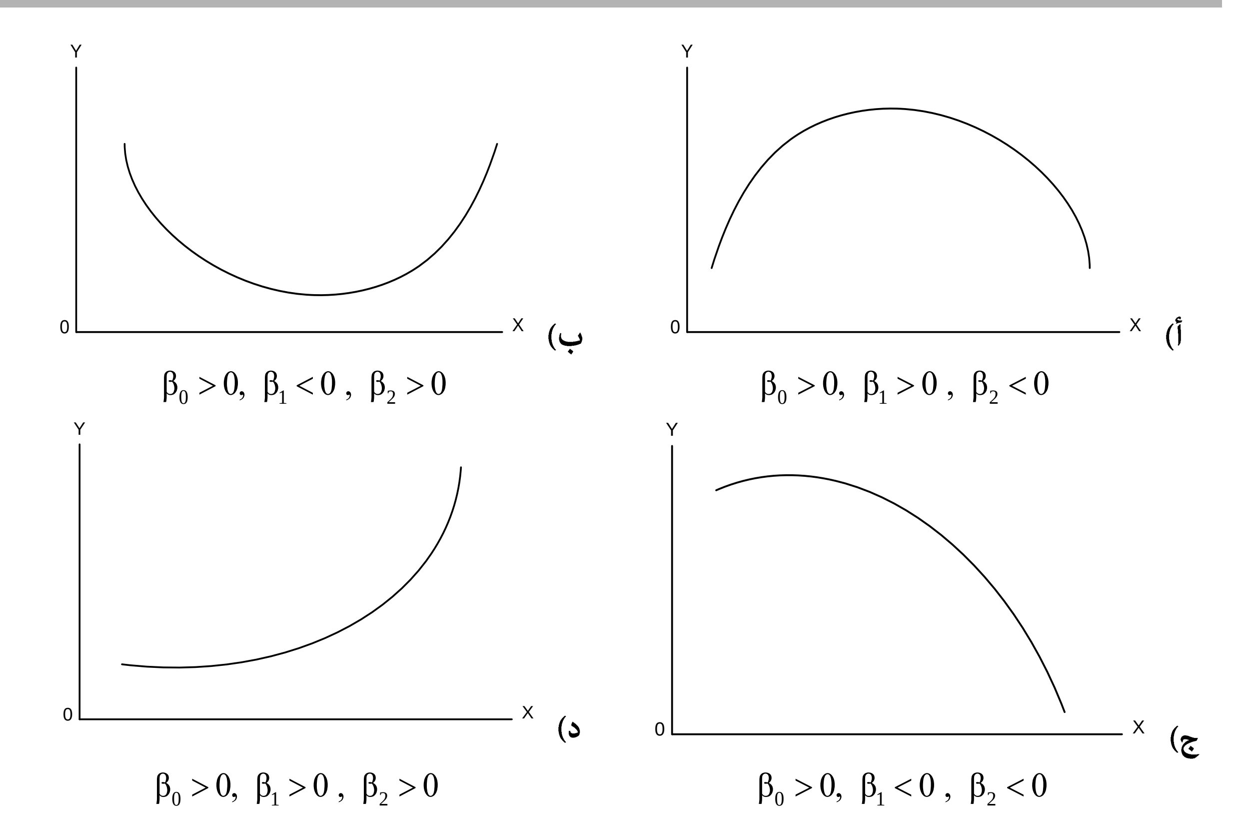
ويستخدم نموذج الانحدار من الدرجة الثانية في الحالات التالية:

- عندما تكون دالة المتغير التابع الحقيقية دالة من الدرجة الثانية يضم مكوني الأثر الخطي والتربيعي معاً.
- عندما تكون دالة الانحدار مجهولة أو معقدة والمعادلة من الدرجة الثانية تمثل أفضل تقدير للدالة المجهولة.

وتوضح الأشكال رقم (٣-١٧-أ)، (٣-١٧-ب)، (٣-١٧-ج) و(٣-١٧-د) حالات مختلفة لمنحنى الانحدار من الدرجة الثانية، حيث يلاحظ أن ميل الدالة غير ثابت بعكس ميل نموذج الانحدار الخطي الذي يتصف بالثبات.

---

* انظر الفصل السابع للتعرف على مشكلة الارتباط الخطي المتعدد والنتائج المترتبة عليها.



شكل رقم (٣-١٧): حالات مختلفة من منحنى نموذج الانحدار من الدرجة الثانية

ب- اختبار الفروض:

توجد ثلاثة أسئلة استدلالية أساسية مرتبطة بنموذج الانحدار من الدرجة الثانية هي:

- هل الانحدار ككل معنوي؟
- هل نموذج الانحدار من الدرجة الثانية يعطي قوة تفسيرية أو تنبؤية أكبر من تلك التي يمكن الحصول عليها من النموذج الخط البسيط؟
- إذا كان نموذج الانحدار من الدرجة الثانية ملائماً هل يمكن إضافة حد آخر (X^3) للنموذج؟

ج- اختبار معنوية الانحدار ككل:

لاختبار معنوية كل المتغيرات المستقلة يتم اختبار الفرض التالي:

فرض العدم: $H_0{:}\beta_0 = \beta_1 = \beta_2 = 0$ **مقابل الفرض البديل:** "ليست كل قيم المعالم مساوية للصفر"

ويكافئ هذا الاختبار اختبار F الذي نحصل عليه من جدول تحليل التباين، أي:

$$F_o: \frac{ESS/p}{RSS/(n-p-1)} \sim F_{p,n-p-1}$$

حيث إن:

ESS= مجموع مربعات الانحدار لنموذج الانحدار من الدرجة الثانية.

RSS= مجموع مربعات البواقي للنموذج.

p= عدد المتغيرات المستقلة (X وX^2).

n= عدد المشاهدات.

ويمكن الوصول لقرار بشأن معنوية الانحدار ككل كما يلي:

- إذا كانت قيمة F_0 أكبر من قيمة توزيع F الجدولية بدرجتي حرية p و(n-p-1) ومستوى معنوية معين، نرفض فرض العدم ونقبل الفرض البديل ومن ثم فإن كل قيم معلمات النموذج لا تساوي الصفر، أي أن المتغيرين (X وX^2) تأثير معنوي على المتغير التابع.
- أما إذا كانت قيمة F_0 أقل من قيمة F الجدولية بدرجتي حرية p و(n-p-1)، نقبل فرض العدم القائل بتساوي كل معلمات النموذج للصفر، أي الانحدار ليس له معنوية إحصائية بمعنى أن المتغيرين لا يؤثران على المتغير التابع.

د- اختبار معنوية إضافة الحد (X^2) للنموذج:

بعد اختبار معنوية الانحدار ككل، يتم عادة اختبار إضافة الحد (X^2)، أي يتم اختبار معنوية التحسن الذي يحدث في المقدرة التفسيرية لنموذج الانحدار الخطي البسيط نتيجة إضافة الحد التربيعي (X^2). ويتم إجراء هذا الاختبار باستخدام اختبار F الجزئي أو اختبار t كما سبق شرحهما. والفرض المراد اختباره في هذه الحالة هو:

فرض العدم: إن إضافة الحد (X^2) لنموذج الانحدار الخطي لا يحسن المقدرة التفسيرية للنموذج $(H_{0:}\ \beta_2 = 0)$.

مقابل الفرض البديل: إن إضافة الحد (X^2) للنموذج يسهم بمستوى معنوي في تفسير المتغير التابع $(H_{1:}\ \beta_2 \neq 0)$.

اختبار إضافة حد آخر لنموذج الانحدار من الدرجة الثانية:

إذا ما تبين لنا من الاختبار أن الحد (X^2) يسهم بمستوى معنوي في تفسير تباين المتغير التابع، يتم بناء نموذج يضم بالإضافة للمتغيرين X و(X^2) الحد (X^3) ويتم اختبار معنوية المتغير الجديد (X^3) بنفس الطريقة السابقة الذكر. فإذا أوضح الاختبار أن المتغير ليس له دلالة إحصائية نحكم بأن نموذج الانحدار من الدرجة الثانية ملائم لتوفيق البيانات.

هـ- تحديد القيم المثلى:

من الاستخدامات الأساسية لنموذج الانحدار التربيعي هو تحديد القيم المثلى للمتغير التابع والمستقل. ولتحديد قيمة المتغير المستقل (X) التي تحقق أعلى (أدنى) قيمة للمتغير التابع يتم إجراء مفاضلة دالة الانحدار التربيعية بالنسبة لـ X ومساواة الناتج بالصفر كما يلي:

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial x}(\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + e)$$
$$= \hat{\beta}_1 + 2\hat{\beta}_2 x = 0$$

أي:

$$x_{optimal} = \frac{-\hat{\beta}_1}{2\hat{\beta}_2} \qquad (3\text{-}70)$$

وللحصول على القيمة العليا (الدنيا) للمتغير التابع يتم التعويض عن قيمة X في معادلة نموذج الانحدار من الدرجة الثانية بالقيمة المثلى ($X_{optimal}$) كما يلي:

$$\hat{y}_{optimal} = \hat{\beta}_0 + \hat{\beta}_1 . \frac{-\hat{\beta}_1}{2\hat{\beta}_2} + \hat{\beta}_2 . \left(\frac{-\hat{\beta}_1}{2\hat{\beta}_2}\right)^2 = \hat{\beta}_0 - \frac{\hat{\beta}_1^{\,2}}{4\hat{\beta}_2} \qquad (3\text{-}71)$$

د- مثال:

يوضح الجدول رقم (٣-٨) بيانات افتراضية تختص بتجربة صُممت لقياس أثر جرعات مختلفة من سماد النيتروجين على إنتاجية محصول القمح. المطلوب هو بناء نموذج انحدار ملائم لتوفيق البيانات.

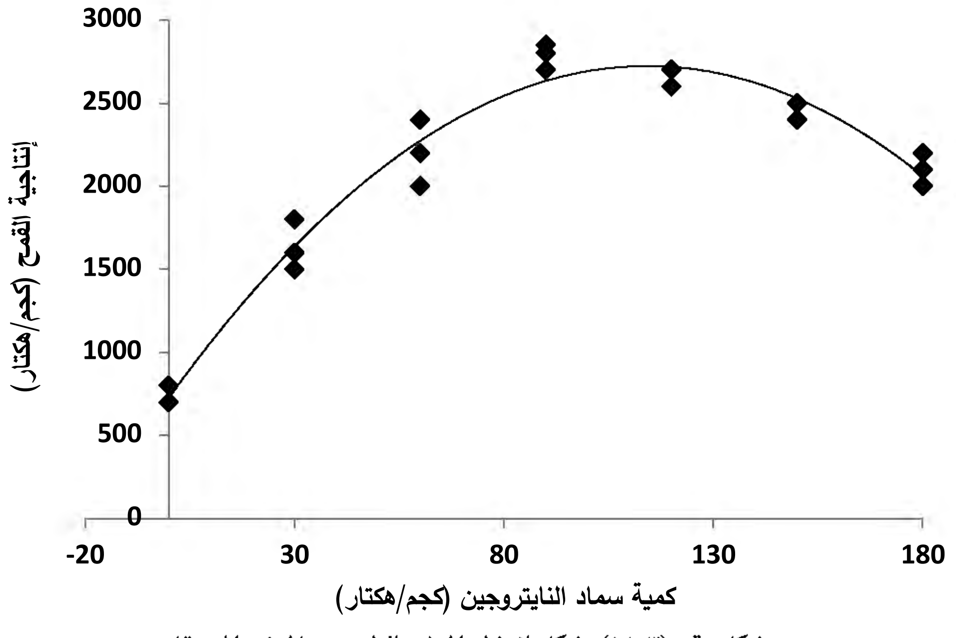
الحل:

للتعرف على العلاقة بين المتغير المستقل (X) والمتغير التابع (Y) نقوم برسم شكل الانتشار الممثل للعلاقة بينهما باستخدام البيانات المعطاة بالجدول رقم (٣-٨). وبمعاينة شكل الانتشار (شكل رقم (٣-١٨)) يتضح أن الصيغة الملائمة لتقدير العلاقة بين المتغير التابع والمتغير المستقل هي الصيغة التربيعية.

جدول رقم (٣-٨): كمية سماد النيتروجين وإنتاجية محصول القمح

رقم المشاهدة	كمية النيتروجين (كجم/هكتار)	إنتاجية القمح (كجم/هكتار)
1	0	700
2	30	1800
3	60	2200
4	90	2700
5	120	2600
6	150	2500
7	180	2200
8	0	800
9	30	1500
10	60	2400
11	90	2800
12	120	2700
13	150	2400
14	180	2100
15	0	700
16	30	1600
17	60	2000
18	90	2850
19	120	2700
20	150	2500
21	180	2000

المصدر: بيانات افتراضية



شكل رقم (٣-١٨): شكل انتشار المتغير التابع مع المتغير المستقل

وباستخدام طريقة المربعات الصغرى الاعتيادية يتم توفيق النموذج التالي:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

حيث إن: y_i = كمية القمح المنتجة (كجم/هكتار)، x_i = كمية النيتروجين المعاملة (كجم/هكتار)، و ε_i= حد الخطأ العشوائي.

وباستخدام الحاسب الآلي (برنامج أكسل) وجد أن نموذج الانحدار الموفق هو:

$$\hat{y}_i = 726.841 + 34.960 x_i - 0.153 x_i^2$$

(0.000) (0.000) (0.000)

F=281.801, Sig. = 0.000

R^2=0.969

S=125.953

حيث تمثل الأرقام الموجودة بين الأقواس أسفل قيم معاملات الانحدار قيم المعنوية أو الدلالة الإحصائية. وتوضح نتائج النموذج الموفق أن كل معاملات الانحدار ذات دلالة إحصائية عالية (p-value<<0.001). ويتضح من ذلك أن المتغيرين (X وX^2) يفسران (٩٧%) تقريباً من التغير في إنتاجية محصول القمح مما يؤكد صحة العلاقة من الدرجة الثانية بين المتغيرين. ولاختبار مدى إسهام إضافة حد آخر(X^3) في تفسير تباين المتغير التابع، تم بناء نموذج الانحدار التالي:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$$

وباستخدام الحاسب الآلي (برنامج اكسل) وجد أن نموذج الانحدار الموفق هو:

$$\hat{y}_i = 715.873 + 36.195 x_i - 0.17152 x_i^2 - 0.000068587 x_i^3$$

(0.000) (0.000) (0.004) (0.720)

F=178.87, Sig.=0.000

R^2=0.969

S=129.099

حيث تمثل الأرقام الموجودة بين الأقواس أسفل قيم معـاملات الانحـدار قـيم المعنويـة أو الدلالـة الإحصـائية. وعـلى الرغم من معنوية الانحدار ككل إلا أننا نجد أن المتغير (X^3) لم يسهم في تفسير تباين المتغير التابع بمستوى معنوي (P-value > 0.05) فضلاً عن أن معامل التحديد (R^2) لم يطرأ عليه أي تحسن يـذكر وأن قيمـة الخطـأ المعيـاري قـد زادت من ١٢٥,٩٥٣ في نموذج الانحدار التربيعي إلى ١٢٩,٠٩٩ بإضافة الحد التكعيبي (X^3). وبالتـالي يمكـن إسـقاط الحـد (X^3) من النموذج ونخلص على أن نموذج الانحدار التربيعي هو النموذج الملائم لتوفيق البيانات.

القيم المثلى:

حدِّد كمية سماد النيتروجين المثلى التي تحقق أعلى إنتاجية من محصول القمح، وما أعلى إنتاجية من محصول القمح يمكن تحقيقها؟

يتم حساب كمية سماد النيتروجين المثلى كما يلي:

$$x_{optimal} = \frac{-\widehat{\beta}_1}{2\widehat{\beta}_2} = \frac{-34.96}{2 \times -0.152998} = 114.25$$

أي أن كمية النيتروجين المثلى التي تحقق أعلى إنتاجية من محصول القمح هي ١١٤,٢٥ كيلوجرام للهكتار. كما يمكن حساب الإنتاجية المثلى التي يمكن تحقيقها كما يلي:

$$\widehat{y}_{optimal} = \widehat{\beta}_0 - \frac{\widehat{\beta}_1^2}{4\widehat{\beta}_2} = 726.98 - \frac{34.96^2}{4 \times -0.152998} = 2724.1$$

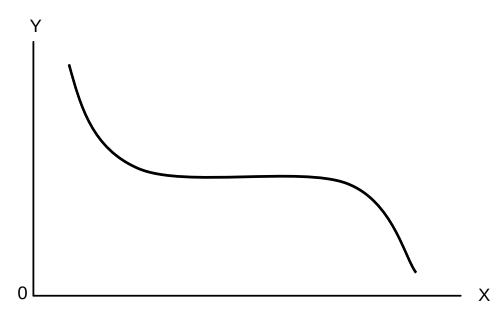
أي أن أعلى إنتاجية يمكن تحقيقها هي ٢٧٢٤,١ كيلوجرام للهكتار.

٣-١٧-٣ نموذج الانحدار من الدرجة الثالثة (Cubic Regression Model):

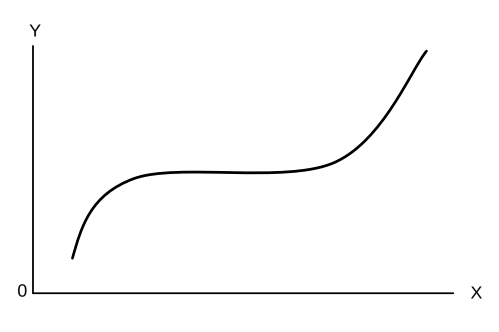
يعتبر نموذج الانحدار من الدرجة الثالثة أو ما يعرف بنموذج الانحدار التكعيبي امتداداً لنموذج الانحدار من الدرجة الثانية. حيث يضم النموذج الحدود (X)، (X^2) و(X^3)، ويصبح شكل العلاقة كالآتي:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$$

ويستخدم نموذج الانحدار من الدرجة الثالثة عندما تكون دالة المتغير الحقيقية دالة من الدرجة الثالثة. ويختلف منحنى معادلة الانحدار من الدرجة الثالثة باختلاف قيم المعالم ($\beta_0\ \beta_1\ \beta_2\ \beta_3$). والشكلان (٣-١٩-أ) و(٣-١٩-ب) يوضحان حالتين من شكل منحنى نموذج الانحدار من الدرجة الثالثة.



الحالة الأولى: $\beta_0 > 0\ ,\ \beta_1 < 0\ ,\ \beta_2 < 0\ ,\ \beta_3 < 0$



الحالة الثانية: $\beta_0 > 0\ ,\ \beta_1 > 0\ ,\ \beta_2 > 0\ ,\ \beta_3 > 0$

شكل رقم (٣-١٩): حالتان مختلفتان من منحني نموذج الانحدار من الدرجة الثالثة

ب- اختبار الفروض:

في نموذج الانحدار من الدرجة الثالثة يتم إجراء الاختبارات التالية:

- اختبار معنوية الانحدار ككل.
- إذا كان الانحدار ككل ذات دلالة إحصائية، يتم اختبار معنوية إضافة الحد X^3 للحدين X وX^2. فإذا أوضح الاختبار أن إضافة الحد X^3 يسهم بمستوى معنوي في تفسير تغير المتغير التابع، نخلص على أن نموذج الانحدار من الدرجة الثالثة ملائم لتوفيق البيانات. ويتم إجراء هذه الاختبارات بنفس الطريقة التي سبق شرحها في حالة النموذج من الدرجة الثانية.

ج- مثال:

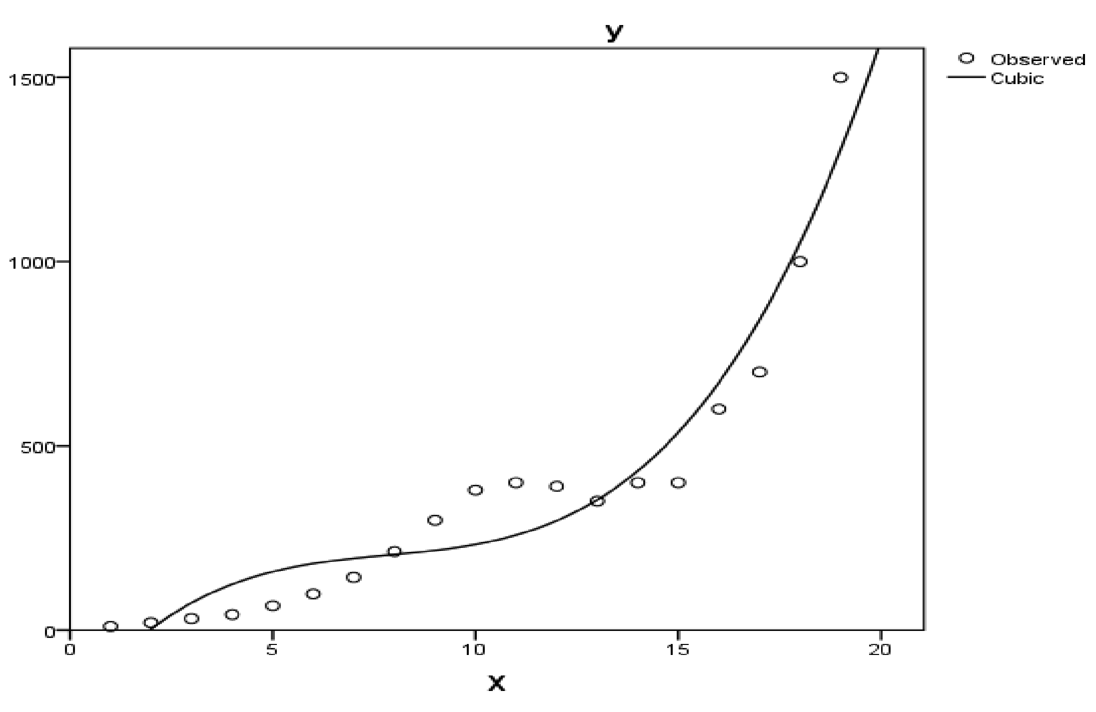
يعطي الجدول رقم (٣-٩) بيانات افتراضية تختص بالتكاليف المتغيرة المقابلة لإنتاج سلعة ما. المطلوب هو بناء نموذج انحدار ملائم لتوفيق البيانات.

جدول رقم (٣-٩): التكاليف المتغيرة والوحدات المنتجة لسلعة ما (بيانات افتراضية).

رقم المشاهدة	الوحدات المنتجة (ألف وحدة)	التكلفة المتغيرة (مليون ريال)
1	1	10
2	2	20
3	3	31
4	4	42
5	5	66
6	6	98
7	7	143
8	8	213
9	9	298
10	10	380
11	11	400
12	12	390
13	13	350
14	14	400
15	15	400
16	16	600
17	17	700
18	18	1000
19	19	1500

الحل:

للتعرف على العلاقة بين المتغير المستقل (X) والمتغير التابع (Y) نقوم برسم شكل الانتشار الممثل للعلاقة بينهما باستخدام البيانات المعطاة بالجدول رقم (٣-٩). وبمعاينة شكل الانتشار (شكل رقم (٣-٢٠)) يتضح أن الصيغة الملائمة لتقدير العلاقة بين المتغير التابع والمتغير المستقل هي الصيغة التكعيبية.



شكل رقم (٣-٢٠): شكل انتشار المتغير التابع والمتغير المستقل

ولإجراء الحل تم أولاً حساب الحدود X^2 وX^3 ومن ثم باستخدام الحاسب الآلي تم توفيق النموذج التالي:

$$\hat{y}_i = -218.41 + 141.11x - 16.71x^2 + 0.71x^3$$

(0.100) (0.017) (0.014) (0.003)

F=65.69, Sig. =(0.000)

R^2=0.929

S=109.9

حيث تمثل الأرقام الموجودة بين الأقواس أسفل قيم معاملات الانحدار قيم المعنوية أو الدلالة الإحصائية. وتوضح نتائج النموذج الموفق أن كل معاملات الانحدار ذات دلالة إحصائية عالية(P-value<<0.001). ويتضح من ذلك أن المتغيرات (x ، x^2 وx^3) تفسر حوالي (٩٣%) من التغير في التكلفة المتغيرة مما يؤكد صحة العلاقة من الدرجة الثالثة بين المتغيرين التابع والمستقل.

ملاحظات:

- من الممكن أن يضم نموذج الانحدار أكثر من متغير مستقل متعدد الحدود، مثال ذلك النموذج التالي:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 z + \beta_4 z^2 + \varepsilon_i$$

حيث يضم النموذج متغيرين من الدرجة الثانية.

- لتجنب مشكلة الارتباط الخطي المتعدد في نماذج الانحدار متعدد الحدود يفضل أن يتم توسيط للمتغيرات المستقلة والتابع وذلك بطرح الوسط الحسابي للمتغير من قيمة أية مشاهدة للمتغير، أي أن يتم التحويل التالي قبل إجراء الانحدار:

$$y'_i = y_i - \bar{y}, \qquad x'_{ji} = x_{ji} - \bar{x}_j \qquad j = 1,2,\ldots,p$$

- من الحلول المقترحة لتجنب مشكلة الارتباط الخطي المتعدد في نماذج الانحدار من الدرجة الثانية فما فوق استخدام ما يعرف بالحدود المتعددة المتعامدة (Orthogonal Polynomials) الأسلوب الذي يتم عن طريقه تحويل المتغيرات المستقلة (الحدود) إلى متغيرات متعامدة (للمزيد حول هذا الأسلوب يرجى الرجوع إلى دريبر وسمث Draper & Smith , 1998, pp461-472).

٣-١٨ اختبار نقص المطابقة (Lack of Fit Test):

تتضمن اشتراطات تحليل الانحدار الخطي أن تكون العلاقة بين المتغير التابع والمتغيرات المستقلة علاقة خطية. ولاستيفاء اشتراط خطية العلاقة، يتم أولاً رسم انتشار المتغير التابع مع كل متغير من المتغيرات المستقلة قبل بناء نموذج الانحدار. ومن ثم يُحدد بالنظر ما إذا كانت نقاط الشكل تقع على امتداد خط مستقيم مائل عن المستوى الأفقي أم لا. فإذا بدا لنا من الشكل أن العلاقة خطية، يتم بناء النموذج وتقييمه من خلال اختبار الدلالة الإحصائية الكلية للنموذج "اختبارF" واختبار t للمتغيرات المستقلة ومعامل التحديد ورسم انتشار بواقي النموذج مع القيم الموفقة و/أو مع أحد المتغيرات المستقلة لتأكيد خطية العلاقة. غير أن رسوم الانتشار ونتائج اختبارات F وt ومعامل التحديد غير كافية للتحقق من استيفاء النموذج الموفق لشرط خطية العلاقة في وجود **تكرار لقيم المتغير المستقل**، خاصة في بحوث تصميم التجارب التي يتم فيها تكرار المعالجات. ففي حالة تكرار قيم المتغير المستقل يتم إضافة اختبار آخر يُعرف باختبار نقص المطابقة (Lack of fit) يتم فيه تجزئة مجموع مربعات البواقي إلى مجموع مربعات الخطأ الصافي (Pure error) ومجموع مربعات عدم المطابقة.

ويتطلب اختبار نقص المطابقة وجود مشاهدات متكررة لقيمة واحدة على الأقل من قيم المتغير المستقل وأن يكون التكرار حقيقي وليس مجرد تكرار قياسات (Montgomery et al., 2012). كما يتطلب الاختبار أن تكون قيم مشاهدات المتغير التابع المقابلة لأي مشاهدة محددة للمتغير المستقل مستقلة بعضها عن بعض وتتبع التوزيع الطبيعي ولها تباين مساو لتباينات مشاهدات المتغير التابع المقابلة لقيم أخرى للمتغير التابع (Montgomery et al., 2012; Weisberg, 2005).

وبفرض أن لدينا n_i مشاهدة من المتغير التابع مقابلة لقيمة محددة للمتغير المستقل x_i $(i=1,2,...,m)$. وأن y_{ij} هي قيمة المشاهدة رقم j المقابلة لقيمة المتغير المستقل رقم i، حيث $(j=1,2,...,n_i)$ ، ويكون مجموع المشاهدات هو $n=\sum_{i=1}^{m} n_i$. وفي هذه الحالة يكون شكل البيانات لنموذج الانحدار الخطي البسيط في حالة وجود تكرار في قيم المتغير المستقل كالتالي:

$$\begin{array}{lllll} x_1 & y_{11} & y_{12} & y_{13} & \cdots & y_{1n_1} \\ x_2 & y_{21} & y_{22} & y_{23} & \cdots & y_{2n_2} \\ x_3 & y_{31} & y_{32} & y_{33} & \cdots & y_{3n_3} \\ . & . & . & . & & . \\ x_m & y_{m1} & y_{m2} & y_{m3} & \cdots & y_{mn_m} \end{array}$$

ولإجراء اختبار نقص المطابقة، يتم تجزئة مجموع مربعات البواقي (RSS) كما يلي (Montgomery et al. 2012):

$$RSS=SS_{PE}+SS_{LOF}$$

حيث إن SS_{PE} مجموع مربعات الخطأ الصافي و SS_{LOF} ومجموع مربعات نقص المطابقة. وتتم التجزئة كما يلي:

$$y_{ij}-\hat{y}_i=\left(y_{ij}-\bar{y}_i\right)+\left(\bar{y}_i-\hat{y}_i\right)$$

حيث إن $\bar{y}_i$ الوسط الحسابي لقيم n_i مشاهدة للمتغير التابع المقابلة لقيمة المتغير المستقل x_i. وبتربيع المعادلة أعلاه والجمع عند جميع مستويات i وj وبعد التبسيط بإجراء بعض العمليات الحسابية نحصل على التالي:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\hat{y}_i\right)^2=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2+\sum_{i=1}^{m}n_i\left(\bar{y}_i-\hat{y}_i\right)^2 \qquad (3\text{-}72)$$

ويتكون مجموع مربعات البواقي من جزأين، الأول مجموع مربعات الخطأ الصافي:

$$SS_{PE}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2$$

وهو مجموع المربعات المصحح للمشاهدات المتكررة عند كل قيمة محددة من قيم المتغير المستقل. فإذا كان فرض ثبات التباين متحققاً فإن SS_{PE} يعتبر مقياسًا للخطأ الصافي ولا يعتمد على النموذج لأن الاختلاف في قيم y عند كل قيمة من x هو الاختلاف الوحيد الذي يستخدم في حساب SS_{PE}. ولوجود (n_i-1) درجة حرية للخطأ الصافي عند كل قيمة x_i فإن العدد الكلي من درجات الحرية المرتبطة بمجموع مربعات الخطأ الصافي هو:

$$\sum_{i=1}^{m}\left(n_i-1\right)=n-m$$

والجزء الثاني هو مجموع مربعات نقص المطابقة هو:

$$SS_{LOF}=\sum_{i=1}^{m} n_i\left(\bar{y}_i-\hat{y}_i\right)^2$$

والذي يمثل مجموع الانحرافات المربعة المرجحة بين $\bar{y}_i$ عند كل قيمة x_i والقيمة المقدرة المقابلة. وعندما تقترب القيم المقدرة $\hat{y}_i$ من $\bar{y}_i$ المقابلة لها فإن ذلك يشير إلى أن نموذج الانحدار نموذج خطي. ولمجموع مربعات نقص المطابقة درجات حرية $(m-2)$ وذلك لوجود m قيمة للمتغير المستقل ودرجتي حرية تم فقدهما نتيجة تقدير معلمتي النموذج (β_0,β_1) للحصول على $\hat{y}_i$. ويتم اختبار نقص المطابقة بحساب الإحصائية التالية:

$$F_0=\frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)}=\frac{MS_{LOF}}{MS_{PE}} \qquad (3\text{-}73)$$

ويلاحظ من المعادلة (3.73) أن القيمة المتوقعة لمتوسط مجموع مربعات الخطأ الصافي MS_{PE} هو σ^2 بغض النظر عن العلاقة الدالية بين المتغيرين التابع والمستقل. والقيمة المتوقعة لـ MS_{LOF} هي:

$$E\left(MS_{LOF}\right)=\sigma^2+\frac{\sum_{i=1}^{m} n_i\left[E(y_i)-\beta_0-\beta_1 x_i\right]^2}{m-2}$$

وتكون القيمة المتوقعة لمتوسط مربعات نقص المطابقة مساو للتباين σ^2 فقط في حالة خطية نموذج الانحدار، أي في حالة $E\left(y_i\right)=\beta_0+\beta_1 x_i$ وبذلك يكون الحد الثاني للمعادلة أعلاه مساوياً للصفر. وتكون القيمة المتوقعة لعدم المطابقة أكبر من التباين $\left\{E\left(MS_{LOF}\right)>\sigma^2\right\}$ في حالة عدم خطية النموذج، أي في حالة $E\left(y_i\right)\neq\beta_0+\beta_1 x_i$. وفي حالة كانت العلاقة خطية بين المتغير التابع Y والمتغير المستقل X في المجتمع، تتبع الإحصاءة F_0 توزيع F بدرجتي حرية (n-2) و(n-m)، أي أن:

$$F_0=\frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)}=\frac{MS_{LOF}}{MS_{PE}}\sim F_{(m-2),(n-m)}$$

ويوضح الجدول رقم (٣-١٠) جدول تحليل التباين لاختبار معنوية نقص المطابقة أو اختبار الفرضية التالية:

$$H_0:E\left(y\right)=\beta_0+\beta_1 x \quad \text{مقابل} \quad H_1:E\left(y\right)\neq\beta_0+\beta_1 x$$

جدول (٣-١٠): جدول تحليل التباين لاختبار معنوية نقص المطابقة

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	قيمة F
الانحدار	1	$ESS=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\hat{y}_{ij}-\bar{y}\right)^2$	$MESS=\frac{ESS}{1}$	$F=\frac{MESS}{MRSS}$
البواقي	n-2	$RSS=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\hat{y}_i\right)^2$	$MRSS=\frac{RSS}{n-2}$	
نقص المطابقة	m-2	$SS_{LOF}=\sum_{i=1}^{m}n_i\left(\bar{y}_i-\hat{y}_i\right)^2$	$MS_{LOF}=\frac{SS_{LOF}}{m-2}$	$F=\frac{MS_{LOF}}{MS_{PE}}$
الخطأ الصافي	n-m	$SS_{PE}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2$	$MS_{PE}=\frac{SS_{PE}}{n-m}$	
المجموع	n-1	$TSS=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}\right)^2$		

وتشير قيمة إحصائية اختبار نقص المطابقة $F=\frac{MS_{LOF}}{MS_{PE}}$ القريبة من الواحد الصحيح إلى خطية العلاقة بين المتغيرين التابع والمستقل، وأما إذا كانت القيمة كبيرة فتشير إلى عدم خطية العلاقة. وفي حال عدم خطية العلاقة بين المتغير التابع والمستقل، يتم إجراء تحويلة للمتغير التابع او للمتغير المستقل أو للمتغيرين معاً أو إضافة حد تربيع إذا ما بدا من شكل الانتشار أن العلاقة تربيعية أو إضافة حدي تربيع x^2 وتكعيب x^3 إذا أظهر شكل الانتشار أن العلاقة تكعيبية بين المتغيرين التابع والمستقل (انظر الفصل الثاني ٢-١٢).

مثال:

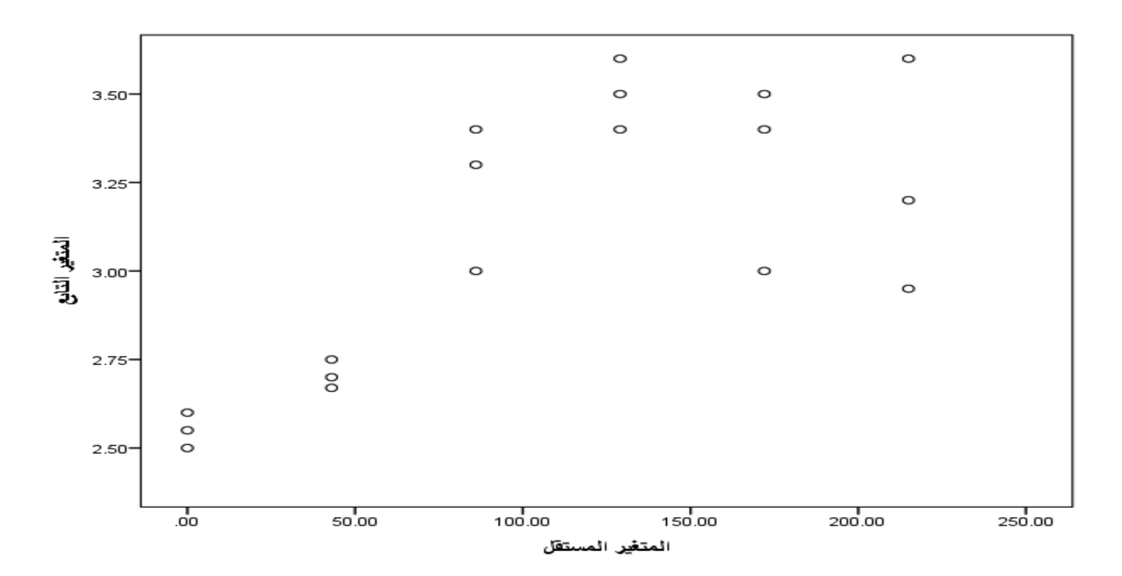
الجدول التالي يوضح بيانات افتراضية لمتغيرين أحدهما تـابع والآخـر مسـتقل. ولأن قـيم المتغـير المسـتقل متكـررة، فالمطلوب استخدام عدم المطابقة للتحقق من خطية العلاقة وتحديد العلاقة الدالية بين المتغيرين في حال عـدم خطيـة العلاقة.

جدول (٣-١١): بيانات افتراضية لمتغيرين أحدهما تابع Y وآخر مستقل X

x_i		y_{ij}	
0.0	2.50	2.60	2.55
43.0	2.70	2.75	2.67
86.0	3.00	3.40	3.30
129.0	3.50	3.60	3.40
172.0	3.40	3.50	3.00
215.0	2.95	3.20	3.60

الحل:

للتعرف على نوع العلاقة بين المتغيرين التابع والمستقل تم أولاً رسم انتشار بين المتغيرين (الشكل رقم ٢١ و٢٢). ويتضح من الشكل أن العلاقة بين المتغيرين علاقة خطية.



شكل رقم (٣-٢١): شكل انتشار المتغيرين التابع والمستقل

وبناءَ على شكل انتشار العلاقة بين المتغيرين التابع والمستقل، تم بناء نموذج انحدار خطي بسيط بين المتغيرين وإجراء اختبار نقص المطابقة لوجود قيم متكررة للمتغير المستقل. ويتضح من نتائج الجدولين التاليين (٣-١٢ و٣-١٣) أن المتغير المستقل له تأثير دال إحصائياً في المتغير التابع ($F_{1,16}$=16.98, p-value=0.001). ولإجراء اختبار عدم المطابقة، تم حساب مجموع مربعات الخطأ الصافي كما يوضح الجدول رقم (٣-١٤). فمثلاً مجموع مربعات الخطأ الصافي لمكررات قيم المتغير التابع عند $x_1 = 0$، يتم حسابه كالتالي:

$$\sum_{j=1}^{3}\left(y_{1j}-\bar{y}_1\right)^2=\left(2.5-2.55\right)^2+\left(2.6-2.55\right)^2+\left(2.55-2.55\right)^2=0.005$$

وهكذا يتم حساب مجموع مربعات الخطأ الصافي لمكررات قيم المتغير التابع المقابلة لبقية قيم المتغير المستقل. ومن ثم يتم جمع المجاميع الفرعية للحصول على مجموع الخطأ الصافي الكلي البالغ (٠,٤٦٩٩)، ويتم الحصول على مجموع مربعات نقص المطابقة بطرح مجموع الخطأ الصافي من مجموع مربعات البواقي.

وبناءً على نتائج اختبار نقص المطابقة ($F_{4,12}$=4.93, p-value=0.014) نرفض فرض العدم $H_0:E(y)=\beta_0+\beta_1x$ وذلك عند مستوى معنوية (٠,٠٥)، وبذلك تكون العلاقة بين المتغيرين غير خطية $H_1:E(y)\neq\beta_0+\beta_1x$.

جدول (٣-١٢): جدول تحليل التباين لنموذج انحدار المتغير التابع على المتغير المستقل

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	قيمة F	قيم الاحتمال p-value
الانحدار	1	1.3185	1.3185	16.98	0.001
البواقي	16	1.2421	0.0776		
نقص المطابقة	4	0.7721	0.193	4.93	0.014
الخطأ الصافي	12	0.4699	0.0392		
المجموع	17	2.5606			

جدول (٣-١٣): نتائج نموذج انحدار المتغير التابع على المتغير المستقل

المتغير	المعامل	الخطأ المعياري	قيمة T	قيم الاحتمال p-value
الثابت	2.6938	0.1164	23.1400	0.0000
المتغير المستقل x	0.0037	0.0009	4.1200	0.0010
الانحراف المعياري= 0.278621، معامل التحديد = 51.5%، معامل التحديد المعدل= 48.5%				

جدول (٣-١٤): حساب مجموع مربعات الخطأ الصافي

قيم x	$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2$	درجات الحرية
0.0	0.005	2
43.0	0.003	2
86.0	0.087	2
129.0	0.020	2
172.0	0.140	2
215.0	0.215	2
المجموع	0.4699	12

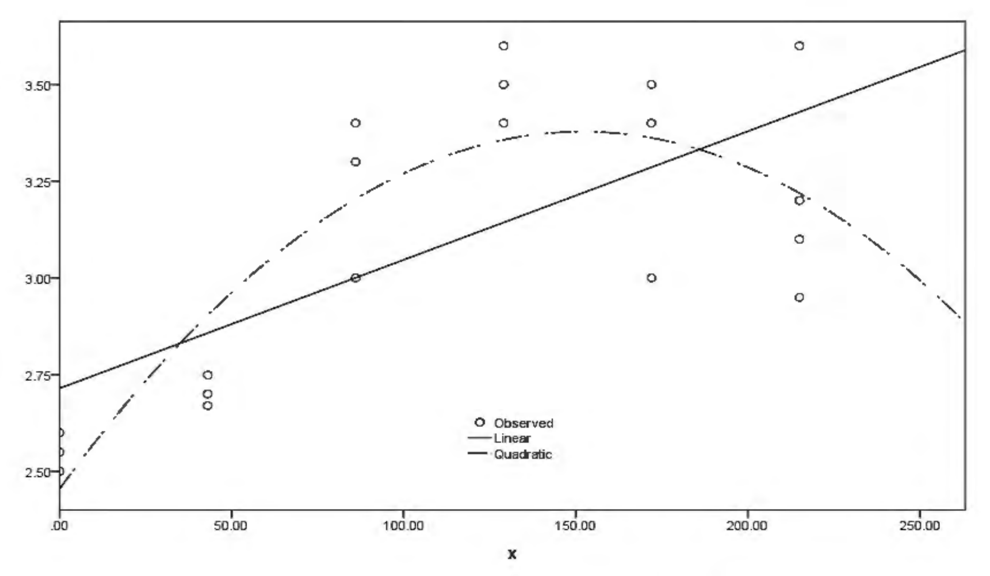
وبناءً على نتيجة عدم خطية العلاقة بين المتغير التابع والمستقل، تم بناء عدد من النماذج الخطية بتحويل المتغير التابع إلى لوغاريتم $\ln(y)$ والجذر التربيعي $\sqrt{y}$ وإضافة حد تربيع المتغير المستقل x^2. وتبين من النتائج أن النموذج الذي يضم حد التربيع هو الأفضل، إذ إنه النموذج الوحيد الذي أوضحت نتائج اختبار نقص المطابقة أن العلاقة بين المتغير التابع والمتغير المستقلين x وx^2 خطية؛ فضلاً عن أن الشكل رقم (٣-٢٢) يشير إلى أن العلاقة بين المتغير التابع والمتغير المستقل علاقة تربيعية (Quadratic). ويوضح الجدولان رقم (٣-١٥) و(٣-١٦) نتائج نموذج الانحدار من الدرجة الثانية. وتشير نتائج اختبار عدم المطابقة $(F_{3,12}=1.85, \text{p-value}=0.191)$ إلى أنه لا يوجد دليل كاف لرفض فرض العدم عند مستوى معنوية (٠,٠٥)، أي أن النموذج $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x+\hat{\beta}_2x^2$ يصف العلاقة بين المتغيرين التابع والمستقل بصورة جيدة.

جدول (٣-١٥): جدول تحليل التباين لنموذج انحدار المتغير التابع على المتغير المستقل

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	قيمة F	قيم الاحتمال p-value
الانحدار	1	1.87294	0.93647	20.43	0.000
البواقي	15	0.68766	0.04584		
نقص المطابقة	3	0.21773	0.07258	1.85	0.191
الخطأ الصافي	12	0.46993	0.03916		
المجموع	17	2.5606			

جدول (٣-١٦): نتائج نموذج انحدار المتغير التابع على المتغير المستقل (دالة من الدرجة الثانية)

المتغير	المعامل	الخطأ المعياري	قيمة T	قيم الاحتمال p-value
الثابت	2.4593	0.1120	21.95	0.0000
المتغير المستقل x	-0.00003805	0.00001094	-3.48	0.003
مربع المتغير المستقل x^2	0.011867	0.002451	4.84	0.0000
الانحراف المعياري= 0.214113، معامل التحديد = 73.1%، معامل التحديد المعدل= 69.6%				



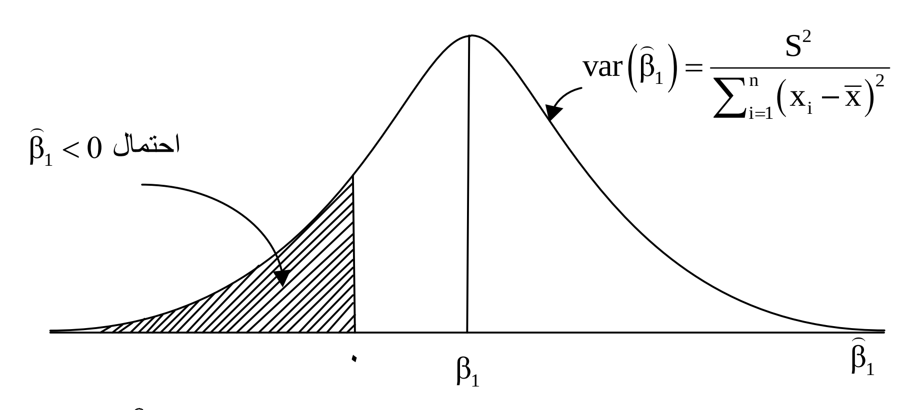
شكل رقم (٣-٢٢): شكل الانتشار بين المتغيرين التابع والمستقل وتمثيل خطي لنموذجي الانحدار الخطي البسيط وانحدار الدرجة الثانية الموفقين

٣-١٩ لماذا تأخذ معاملات نموذج الانحدار الخطي إشارات خاطئة؟

عند بناء نموذج الانحدار المتعدد نحصل أحياناً **على** إشارات معاملات الانحدار الجزئية لبعض المتغيرات المستقلة تخالف النظرية العلمية أو نتائج البحوث السابقة والملاحظة أو الخبرة حول موضوع البحث. فمثلاً نعلم من النظرية الاقتصادية أن هناك علاقة طردية بين الادخار وسعر الفائدة وكذلك بين الاستهلاك والدخل. فحصول إشارة سالبة لمعامل سعر الفائدة في نموذج انحدار الادخار **على** سعر الفائدة مثلاً يخالف النظرية الاقتصادية، مما يصعب تفسير نتائج النموذج الموفق. فالسؤال لماذا نحصل **على** إشارات خاطئة؟ فيما يلي بعض الأسباب الرئيسة للحصول على إشارات خاطئة:

١. ضيق مدى قيم مشاهدات المتغير المستقل:

يعد تقارب قيم المتغير المستقل من أهم أسباب الحصول على إشارة خاطئة. ففي نموذج الانحدار الخطي البسيط، تصبح العلاقة بين تباين معامل الانحدار $\left\{ \text{var}\left(\hat{\beta}_1\right) = \frac{S^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \right\}$ مع تشتت قيم مشاهدات المتغير المستقل علاقة عكسية، فكلما كانت قيم مشاهدات المتغير المستقل قريبة لبعضها البعض كان تباين معامل الانحدار كبيراً. كما أن كبر تباين معامل الانحدار الخطي قد يؤثر في بعض الحالات في تغيير إشارة المعامل (Montgomery, Peck and Vining, 2001, p.120). كما يسهم صغر حجم العينة (عدد المشاهدات) في كبر تباين معامل الانحدار. ويتضح من توزيع المعاينة لمعامل الانحدار المقدر $\hat{\beta}_1$ (الشكل رقم ٣-٢٣) أن احتمال الحصول على تقدير سالب لمعامل الانحدار يعتمد على مدى قرب المعامل للصفر وتباين المعامل الذي يتأثر بمدى تشتت قيم مشاهدات المتغير المستقل.



شكل رقم (٣-٢٣): توزيع المعاينة لمعامل الانحدار المقدر $\hat{\beta}_1$

المصدر: Montgomery, et al. 2001, p.120

٢. عدم إدخال بعض المتغيرات المستقلة في نموذج الانحدار:

من أهم اشتراطات نموذج الانحدار الخطي أن يكون توصيف النموذج صحيحاً. والنموذج الصحيح يتضمن جميع المتغيرات المستقلة المهمة. فإسقاط متغير واحد مهم أو أكثر يؤدي أحياناً إلى الحصول على إشارة خاطئة لمعامل الانحدار. ولتوضيح أثر إسقاط متغير مستقل المتغير، أورد (Montgomery, et al., 2001, p.121) بيانات نموذج انحدار خطي لمتغيرين مستقلين (X_1, X_2) كما يوضح الجدول رقم (٣-١٧).

جدول (٣-١٧): بيانات عن متغير تابع ومتغيرين مستقلين

المتغير التابع Y	المتغير المستقل الأول X_1	المتغير المستقل الثاني X_2
١	٢	١
٥	٤	٢
٣	٥	٢
٨	٦	٤
٥	٨	٤
٣	١٠	٤
١٠	١١	٦
٧	١٣	٦

المصدر: Montgomery et al. 2012

وببناء نموذج الانحدار الخطي البسيط بإدخال المتغير المستقل الأول (X_1) فقط نحصل على النموذج الموفق التالي:

$$\hat{y} = 1.835 + 0.4631x_1$$

وببناء نموذج الانحدار الخطي المتعدد بإدخال المتغيرين المستقلين (X_1, X_2) نحصل على النموذج الموفق التالي:

$$\hat{y} = 1.0355 - 1.2223x_1 + 3.6493x_2$$

ويلاحظ من نتيجة النموذجين أن إشارة معامل المتغير المستقل الأول (X_1) موجبة في النموذج الأول وسالبة في النموذج الثاني، مما يشير إلى أهمية إدخال جميع المتغيرات المستقلة التي تؤثر في المتغير التابع.

٣. مشكلة الارتباط الخطي المتعدد للمتغيرات المستقلة (multicollinearity):

من الاشتراطات الأساسية التي تقوم عليها طريقة المربعات الصغرى هو عدم وجود علاقة خطية تربط بين أحد المتغيرات المستقلة ومتغير مستقل آخر أو بين أحد المتغيرات المستقلة وأية تركيب خطي بين المتغيرات المستقلة الأخرى. ومن أهم المؤشرات التي تدل على وجود مشكلة الارتباط الخطي المتعدد هو أن يكون معظم أو كل معاملات الانحدار الجزئية غير دالة إحصائياً على الرغم من كبر حجم معامل التحديد والدلالة الإحصائية الكلية للانحدار. كما أن الإشارات الخاطئة لمعاملات الانحدار الجزئية قد تعزى إلى وجود مشكلة الارتباط الخطي المتعدد ذلك لأن وجود الارتباط الخطي المتعدد يزيد من قيم تباينات معاملات المتغيرات المستقلة مما يسهم في احتمال الحصول على إشارات خاطئة.

٤. وجود مشاهدات شاذة (Outliers):

نواجه أحياناً في تحليل الانحدار بوجود عدد قليل من القيم الشاذة في مشاهدات المتغير التابع أو المتغيرات المستقلة أو في المتغير التابع والمتغيرات المستقلة معاً. والمشاهدات الشاذة هي مجموعة قليلة من المشاهدات تبعد قيمها بصورة كبيرة عن بقية قيم المشاهدات في العينة. ويرجع ظهور البيانات الشاذة إلى أخطاء إما في مرحلة جمع البيانات أو في مرحلة المعالجة كإدخال البيانات في الحاسب الآلي، وقد تكون هذه البيانات حقيقية ناتجة عن ظروف غير عادية كحدوث كوارث طبيعية، كالزلازل والأعاصير والأمطار الغزيرة، يؤثر على مستويات الإنتاج الزراعي مثلاً.

إن وجود قيم شاذة في مشاهدات المتغير التابع أو المتغيرات المستقلة أو في المتغير التابع والمتغيرات المستقلة معاً يؤثر في تقديرات وإشارات معاملات نموذج الانحدار والإحصاءات المرتبطة بها.

مثال:

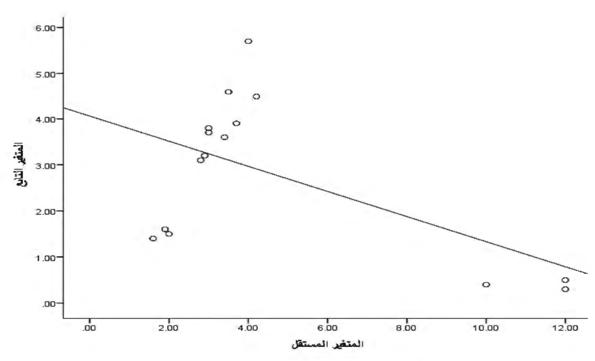
البيانات التالية تشمل قيم شاذة في مشاهدات المتغير المستقل. المطلوب رسم انتشار المتغيرين التابع والمستقل وتحديد المشاهدات الشاذة وبناء نموذجين أحدهما متضمنٌ جميع المشاهدات والآخر باستبعاد المشاهدات الشاذة.

جدول رقم (٣-١٨): بيانات افتراضية عن متغيران تابع ومستقل

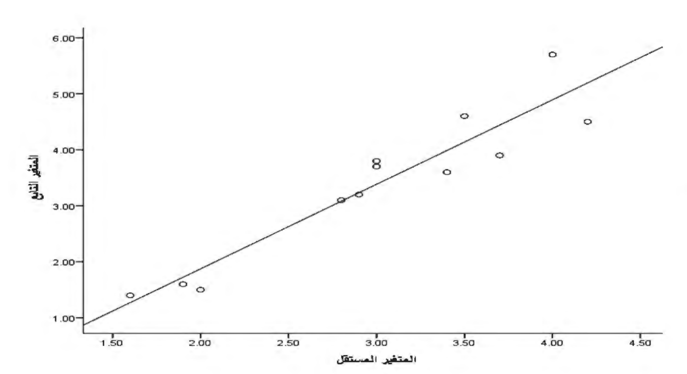
رقم المشاهدة	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢	١٣	١٤	١٥
المتغير التابع Y	5.7	3.8	4.6	3.7	3.9	0.5	4.5	1.5	3.6	0.3	1.4	1.6	3.1	0.4	3.2
المتغير المستقل x	4	3	3.5	3	3.7	12	4.2	2	3.4	12	1.6	1.9	2.8	10	2.9

الحل:

يوضح الشكل رقم (٣-٢٤) رسم انتشار المتغيرين التابع والمستقل. ويلاحظ من الشكل رقم (٣-٢٤-أ) وجود ثلاث قيم في مشاهدات المتغير المستقل، هي: المشاهدة رقم (٦) و(١٠) و(١٤). ويلاحظ من نتائج نموذجي الانحدار (الجدول رقم ٣-١٩)، أن وجود قيم شاذة يؤثر في تقديرات معاملات الانحدار والإحصائيات الأخرى كمعامل التحديد. ويلاحظ أن معامل المتغير المستقل سالب في النموذج الذي يتضمن المشاهدات الشاذة وموجب في النموذج الذي لا يحتوي على المشاهدات الشاذة الثلاث (٦، و١٠، و١٤).



شكل رقم (٣-٢٤-أ): رسم الانتشار لجميع البيانات بما في ذلك القيم الشاذة



شكل رقم (٣-٢٤-ب): رسم الانتشار بعد استبعاد القيم الشاذة

جدول (٣-١٩): نتائج نموذجي الانحدار بوجود وعدم وجود مشاهدات شاذة

	بوجود المشاهدات الشاذة		باستبعاد المشاهدات الشاذة	
	المعامل	مستوى الدلالة	المعامل	مستوى الدلالة
الثابت	4.0597	0.000	-1.1444	0.058
المتغير المستقل	-0.2728	0.027	1.5093	0.000
الدلالة الكلية	0.027		0.000	
معامل التحديد	%32.2		%88.5	

٣-٢٠ تحليل الانحدار باستخدام بعض برامج الإحصاء الجاهزة:

توجد في الوقت الحالي العديد من برامج الإحصاء الجاهزة التي تقوم بجميع أنواع التحليل الإحصائي. وتعد برامج SAS وSPSS وBMDP وMINITAB وSYSTAT وSTATA وStatgraphics من أهم هذه البرامج وأوسعها استخداماً في التحليل الإحصائي. ولا تتطلب هذه البرامج أي معرفة ببرمجة الحاسب بل إنها تعتمد على مجموعات من الجمل والأوامر التي تستخدم لإجراء تحليل إحصائي محدد. وفي هذا الجزء سنقوم بتحليل بيانات مثال انحدار وزن الطفل على طول وعمر الطفل باستخدام برامج SAS، وSPSS، وEXCEL.

٣-٢٠-١ برنامج[§] SAS:

نظام التحليل الإحصائي ساس (**S**tatistical **A**nalysis **S**ystem (SAS)) عبارة عن مجموعة من البرامج الجاهزة تستخدم في عملية التحليل الإحصائي للبيانات. ويتم تحليل البيانات في نظام ساس بإحدى طريقتين، في الطريقة الأولى يتم كتابة برنامج يحدد فيه المتغيرات المراد تحليلها وأداة التحليل الإحصائي من خلال واجهة تطبيق مبنية على نوافذ سهلة الاستخدام، وفي الطريقة الثانية يتم تحليل البيانات من خلال شريط قوائم (Menu-based interface) يتم من خلاله إدخال البيانات أو استيرادها من برامج أخرى كبرنامج إكسل (Excel) وAccess ... إلخ أو قراءتها مباشرة من الإنترنت ومن ثم الاختيار من القائمة نوع التحليل المطلوب والحصول مباشرة على نتائج التحليل.

٣-٢٠-١-١ إجراء تحليل الانحدار من خلال كتابة أوامر وإجراءات:

يتم إجراء التحليل الإحصائي من خلال ثلاث نوافذ، هي نافذة المحرر (Editor) يتم من خلالها تحرير الإجراءات ونافذة المتابعة (log) والتي تظهر والتحذيرات والأخطاء في البرنامج الذي تم تحريره في حال وجودها ونافذة مخرجات التحليل (output) والتي تظهر نتائج التحليل الإحصائي (إطار رقم ٣-٢). ولإجراء أي تحليل إحصائي يتم كتابة برنامج خاص عن طريق نافذة المحرر (Editor) به يحدد فيه البيانات والمتغيرات المراد تحليلها وأداة التحليل الإحصائي. وتنقسم أوامر ساس لإجراء أي تحليل إحصائي إلى جزأين: يتم في الجزء الأول تعريف المتغيرات وإدخال البيانات والذي يبدأ بكلمة DATA التي تعقبها كلمة INPUT لتحديد أسماء المتغيرات. وتستخدم كلمة datalines -التي تلي كلمة INPUT -عندما تكون البيانات المراد معالجتها موجودة في نفس البرنامج كما هو الحال في هذا المثال (إطار رقم (٣-٣)) وتستخدم عبارة INFILE عندما تكون البيانات مخزنة في ملف خارجي. وأما الجزء الثاني فيختص بالإجراءات المطلوبة لإجراء أي تحليل إحصائي، ولكل إجراء اسمه الخاص به لإجراء التحليل المحدد.

[§] للمزيد حول نظام ساس يرجى الرجوع إلى SAS/STAT User's Guide Version 9.3 و Cody & Smith (2005) و Marasinghe and Kennedy (2008)

إطار رقم (٣-٢): نوافذ نظام ساس (اصدار ٩,٣) عند تشغيل البرنامج



إطار رقم (٣-٣): شكل البيانات والأوامر المطلوبة لإجراء تحليل الانحدار باستخدام إجراء PROC REG:

```
DATA CHILD;
  INPUT weight age height;
   datalines;
   11.50     3.00     84.00
   16.00     5.00     95.00
     .         .        .
     .         .        .
     .         .        .
    3.30     0.00     49.00
    1.40     0.08     40.00
;
PROC REG;
  MODEL weight = age height;
    RUN;
```

إجراءات تحليل الانحدار:

يتيح نظام ساس مجموعة كبيرة من الإجراءات يمكن من خلالها إجراء تحليل الانحدار الخطي. وفيما يلي نستعرض بعض الإجراءات التي تستخدم في تحليل الانحدار الخطي:

- **إجراء تحليل الانحدار PROC REG:**

يستخدم الإجراء PROC REG لتقدير نماذج الانحدار الخطية بواسطة طريقة المربعات الصغرى الاعتيادية (OLS). والصيغة العامة لهذا الإجراء هي:

PROC REG;
MODEL *dependents = regressors /options*;

ويتم من خلال هذا الإجراء إجراء نموذج انحدار المتغير/المتغيرات التابعة (dependents) على المتغيرات المستقلة (regressors). فإذا كان هناك أكثر من متغير تابع فإن تحليل الانحدار يُجرى لكل متغير من المتغيرات التابعة على المتغيرات المستقلة على حدة. ويشتمل هذا الإجراء على عدد كبير من الخيارات للحصول على بعض الإحصاءات المرتبطة بنموذج الانحدار المقدر. وفيما يلي بعض الخيارات المهمة المتاحة:

خيارات الإجراء PROC REG *options*;:

- *SIMPLE* : يتيح هذا الخيار طباعة الإحصاءات الوصفية – المجموع (Sum)، الوسط الحسابي (Mean) ومجموع المربعات غير المصحح (Uncorrected SS)، التباين (Variance) والانحراف المعياري (Std Deviation-) لكل المتغيرات المضمنة في النموذج (التابعة والمستقلة).
- *CORR*: يتيح هذا الخيار طباعة مصفوفة معاملات الارتباط الخطي البسيط بين أزواج المتغيرات المستقلة والتابع.
- *USSCP* : يتيح هذا الخيار طباعة مصفوفة مجموع المربعات غير المصحح (Uncorrected Sum of Squares) لكل المتغيرات (المستقلة والتابع).

خيارات النموذج MODEL *dependents = regressors / options*;:

- NOINT : يتيح هذا الخيار تقدير النموذج بدون المعامل الثابت (Intercept).
- XPX : طباعة المصفوفة $\left(\mathbf{X}^{T}\mathbf{X}\right)$.
- I : طباعة المصفوفة $\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}$.
- SS2 : طباعة مجموع المربعات الجزئي (Type II SS) لكل معامل من معاملات النموذج المقدرة.

- STB : طباعة معاملات الانحدار المعياري. ويتم حساب المعامل المعياري بقسمة قيمة المعامل على نسبة الانحراف المعياري للمتغير التابع للانحراف المعياري للمتغير المستقل.
- VIF: طباعة عامل التضخم لكل متغير (انظر الفصل السابع).
- COLLIN: طباعة القيم الكامنة/المميزة مؤشر ورقم الحالة لكل معاملات النموذج بما في ذلك المعامل الثابت. حيث تستخدم هذه المقاييس في الكشف عن وجود مشكلة الارتباط الخطي المتعدد (انظر الفصل السابع).
- COLLINOINT: طباعة القيم الكامنة/المميزة مؤشر ورقم الحالة لكل معاملات النموذج ما عدا المعامل الثابت. حيث تستخدم هذه المقاييس في الكشف عن وجود مشكلة الارتباط الخطي المتعدد (انظر الفصل السابع).
- P: حساب القيم المقدرة/ المتنبأ بها باستخدام قيم البيانات المستخدمة في تقدير النموذج. وتحتوي المخرجات على رقم المشاهدة والقيم الحقيقية للمتغير التابع والقيم المقدرة باستخدام النموذج.
- R : طباعة تحليل البواقي الذي يحتوي على البواقي الاعتيادية، بواقي ستودنت ومقياس كوك (انظر الفصل الرابع).
- Partial: طباعة رسوم الانحدار الجزئية.
- CLM: طباعة فترة ثقة (٩٥%) للقيمة المتوقعة للمتغير التابع لكل مشاهدة.
- CLI: طباعة فترة ثقة (٩٥%) للتنبؤ الفردي (التنبؤ بمشاهدة جديدة).
- DW: طباعة إحصاء ديربن-واتسون (انظر الفصل السابع).
- INFLUENCE: طباعة تحليل تفصيلي عن تأثير كل مشاهدة على مقدرات المعالم وقيم المتغير التابع المقدرة (انظر الفصل الرابع).
- SELECTION= *option*: يعتبر هذا الخيار من أهم خيارات هذا الإجراء؛ إذ يحتوي على طرق اختيار المتغيرات المستقلة في نموذج (انظر الفصل السادس). ويحتوي هذا الخيار على خيارات تشمل طرق الاختيار التالية:
  - FORWARD: طريقة الإضافة إلى الأمام.
  - BACKWARD: طريقة الحذف إلى الخلف.

- STEPWISE[**]: طريقة الانحدار التدرجي.
- MAXR: طريقة اختيار إلى الأمام تتضمن توفيق أفضل نموذج يضم متغير مستقل واحد وأفضل نموذج يضم متغيرين مستقلين وهكذا. ويستخدم معامل التحديد كمعيار للمفاضلة بين النماذج.
- MINR: تشبه طريقة MAXR باستثناء أنه يتم استبعاد المتغيرات المستقلة ذات المساهمة الضعيفة في معامل التحديد.
- RSQUARE[2]: طريقة اختيار عدد محدد من النماذج التي لديها أكبر قيم لمعامل التحديد.
- ADJRSQ: طريقة اختيار عدد محدد من النماذج التي لديها أكبر قيم لمعامل التحديد المعدلة.
- CP: طريقة اختيار عدد محدد من النماذج التي لديها أقل قيم لإحصاء ملاوس.

- **إجراء تحليل الانحدار GLM:**

يستخدم هذا الإجراء طريقة المربعات الصغرى لتوفيق النموذج الخطي العام. وباستخدام هذا الإجراء يمكن إجراء نموذج الانحدار الخطي البسيط والمتعدد، ونماذج الانحدار متعددة الحدود، ونموذج الانحدار المرجح، وتحليل التباين، تحليل التغاير، وتحليل التباين المتعدد المتغيرات وتحليل التباين للقياسات المتكررة. والصيغة العامة لهذا الإجراء لتحليل الانحدار هي:

PROC GLM *options*;
MODEL *dependent* = *independents* /*options*;

- **إجراء تحليل الانحدار[††] RSQUARE**

يستخدم هذا الإجراء للحصول على أفضل نموذج يضم مجموعة من المتغيرات المستقلة بغرض التنبؤ بالمتغير التابع. ويستخدم هذا الإجراء معامل التحديد كمعيار للمفاضلة بين النماذج الممكن توفيقها. والصيغة العامة لهذا الإجراء هي:

PROC RSQUARE *options*;
MODEL *dependents* = *independents* /*options*;

ومن الخيارات المفيدة في هذا الإجراء الخيار B الذي بإضافته نحصل على قيم معاملات الانحدار المقدرة لكل نموذج مُختار.

[**] توجد هذه الخيارات أيضاً كإجراءات مستقلة كما سنرى لاحقاً.

[††] هذا الإجراء مضمن كخيار نموذج في إجراء PROC REG.

- **إجراء تحليل الانحدار [3] STEPWISE:**

يستخدم هذا الإجراء للحصول على أفضل نموذج يضم مجموعة من المتغيرات المستقلة. ويحتوي هـذا الإجـراء عـلى طرق اختيار الانحدار التدرجي، الاختيار إلى الأمام والحذف من الخلف. والصيغة العامة لهذا الإجراء هي:

PROC STEPWISE;
MODEL *dependents = independents /options*;

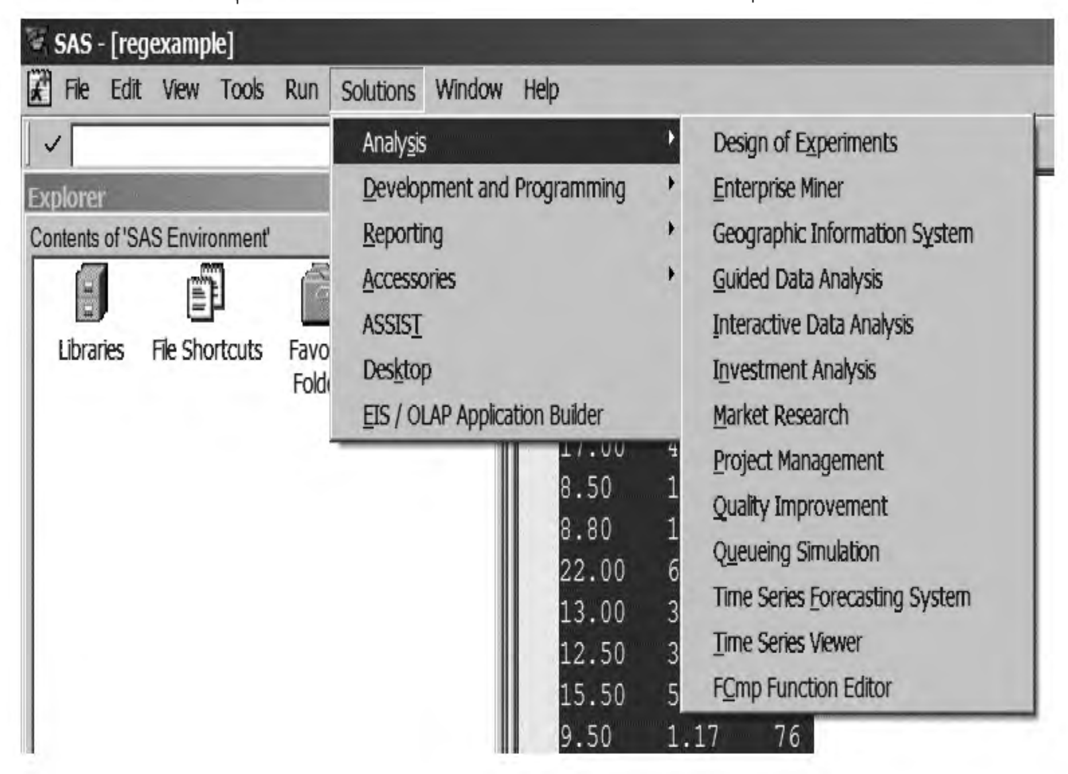
يحتوي خيارات النموذج على التالي:

- NOINT: يتيح هذا الخيار بناء النموذج بدون المعامل الثابت (Intercept).
- FORWARD: طريقة الإضافة إلى الأمام.
- BACKWARD: طريقة الحذف إلى الخلف.
- STEPWISE: طريقة الانحدار التدرجي.
- MAXR: تقوم هذه الطريقة بإيجاد أفضل نموذج يضم متغيراً مستقلاً واحداً الذي يفسر أكبر نسبة مـن تبـاين Y (النموذج الذي لديه أكبر قيمة معامل تحديد). ومن ثـم تـتم إضـافة متغـير آخـر يسـهم بـأكبر زيـادة في قيمـة معامل التحديد. وبمجرد الحصول على النموذج الذي يضم متغيرين، تتمّ مقارنة كل واحد منهما بالمتغيرات التي تتضمن حتى الآن في النموذج. وفي كل مقارنة تحدد طريقة MAXR إذا ما كان حذف أحد المتغيرين واستبداله بآخر يمكن أن يسهم في زيادة معامل التحديد. وفي حالة وجود متغير يمكن أن يسهم في زيادة معامـل التحديـد أكبر من أحد المتغيرات المضمنة في النموذج، يتم حذف الثاني وإدخال الأول. وتسـتمر هـذه العمليـة إلى أن يـتم الوصول إلى أفضل نموذج يضم متغيرين. ومن ثم في الخطوة التالية تتم إضافة متغير ثالث على أساس مساهمته في زيادة معامل التحديد. وتتم مقارنة أي متغير مضمن في النموذج بالمتغيرات التـي لم تتضـمن وهكـذا تسـتمر العملية للوصول إلى أفضل نموذج يضم ثلاثة متغيرات أو أكثر.
- MINR: تشبه هذه الطريقة طريقة MAXR باستثناء أنه في خطوة يتم استبعاد المتغير المستقل ذي المساهمة غـير الدالة إحصائياً في معامل التحديد.
- مستوى الدلالة (المعنوية): يتيح هذا الإجراء خيار تحديد مستوى المعنوية (SLE) لتحديد المتغـير الـذي سـيدخل النموذج لطريقتي الاختيـار إلى الأمـام والانحـدار التـدرجي ومسـتوى المعنويـة (SLS) للمتغـير الـذي سـيبقى في النموذج في طريقتي الانحدار التدرجي والحذف من الخلـف. فـإذا لم يـتم تحديـد مسـتوى المعنويـة، يسـتخدم النظام مستوى المعنويـة (SLE=0.5) لطريقـة الاختيـار للإمـام و(SLE=0.15) لطريقـة الانحـدار التـدرجي ومسـتوى معنوية (0.10) لبقاء المتغـير المسـتقل في طريقـة الاختيـار التـدرجي ومسـتوى المعنويـة (0.15) لبقـاء المتغـير في النموذج في طريقة الحذف من الخلف.

٣-٢٠-١-٢ إجراء تحليل الانحدار من خلال شريط القوائم في نظام ساس:

يتيح نظام ساس لتحليل البيانات باستخدام شريط قوائم (Toolbar) إدخال البيانات أو استيرادها من برامج أخرى ومن ثم الاختيار مباشرة من القائمة نوع التحليل المطلوب والحصول مباشرة على نتائج التحليل. ويوضح الإطار رقم (٣-٤) شريط القوائم الذي يحتوي على عدة خيارات والذي يتم من خلالها إدخال البيانات أو حفظها أو فتح ملف أو استيراد ملف بيانات، وتحديد الأداة الإحصائية لتحليل البيانات من خلال قائمة Solution وAnalysis التي تحوي عدة خيارات للتحليل كتصميم التجارب (Design of experiments) ونظام المعلومات الجغرافية (Geographic Information System) وتحليل البيانات الموجه (Guided data analysis) وغيرها.

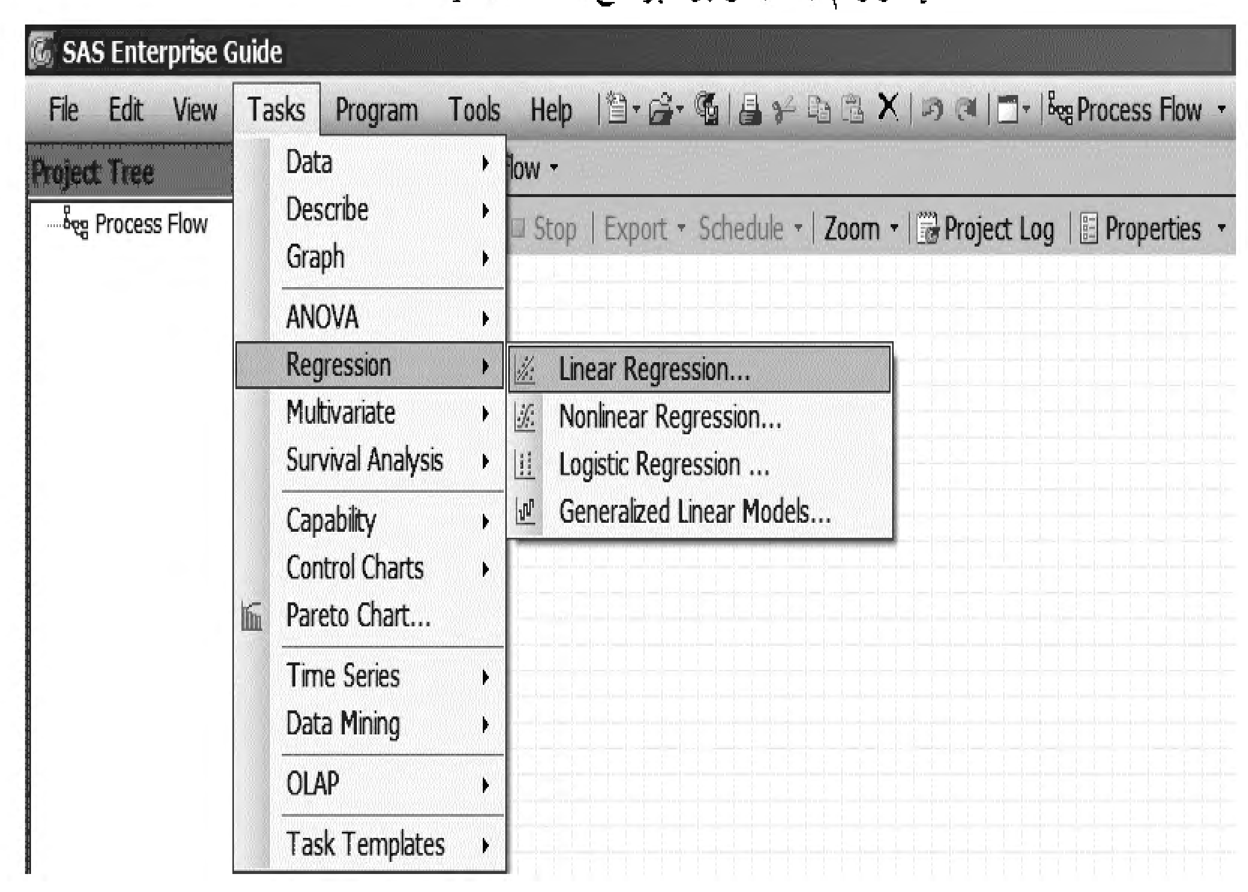
إطار رقم (٣-٤): تحليل البيانات من خلال شريط القوائم



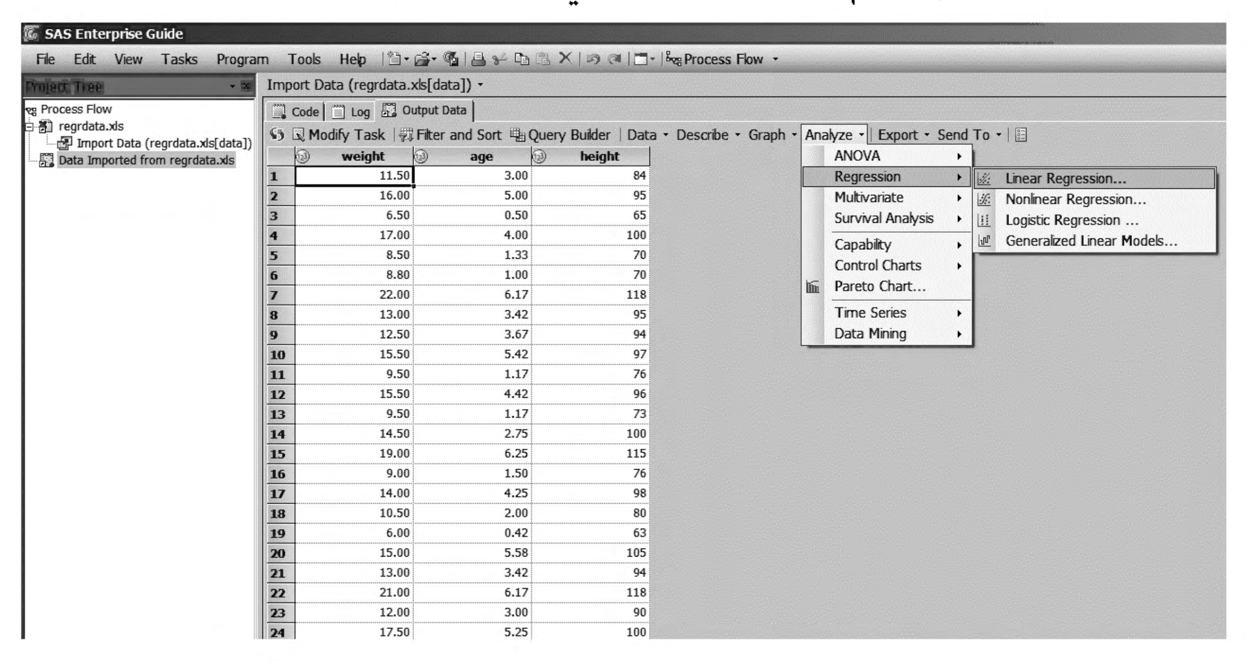
كما طورت شركة ساس برنامج SAS Enterprise Guide والذي يستخدم أشرطة قوائم وواجهات تفاعلية كبرنامج SPSS أو Minitab وغيرهما من البرامج الإحصائية. وللبرنامج شريط قوائم سهل الاستخدام يمكن من خلاله فتح ملف بيانات جديد أو استيراد ملف بيانات أو فتح ملف موجود أصلاً ومن ثم من خلال Tasks أومن ملف البيانات مباشرة من خلال Analyze تحديد الأداة الإحصائية (إطار رقم ٣-٥) والحصول على النتائج مباشرة في نافذة مختلفة. ويتم تحليل الانحدار الخطي باختيار إما Tasks أو Analyze وLinear regression وتحديد المتغير التابع والمتغيرات المستقلة كما يوضح الإطارين رقم (٣-٥) و(٣-٦). وكما يتضح من الإطارين (٣-٧) و(٣-٨) أن البرنامج يتيح خيارات

مختلفة للنموذج كاختيار النموذج الكامل (Full model fitted) أو اختيار طريقة الإضافة إلى الأمام (Forward selection)، ... إلخ (انظر الفصل السادس). كما يمكن من خلال Statistics الحصول على عدد كبير من النتائج من أهمها معاملات الانحدار المعيارية (Standarized regression coefficient) كما يمكن الحصول على قيم مؤشرات فحص النموذج كتحليل الارتباط الخطي المتعدد (Collinearity analysis)، وغيرها وكذلك الحصول على قيم الارتباط الجزئي (Partial correlations).

إطار رقم (٣-٥): واجهة برنامج SAS Enterprise Guide

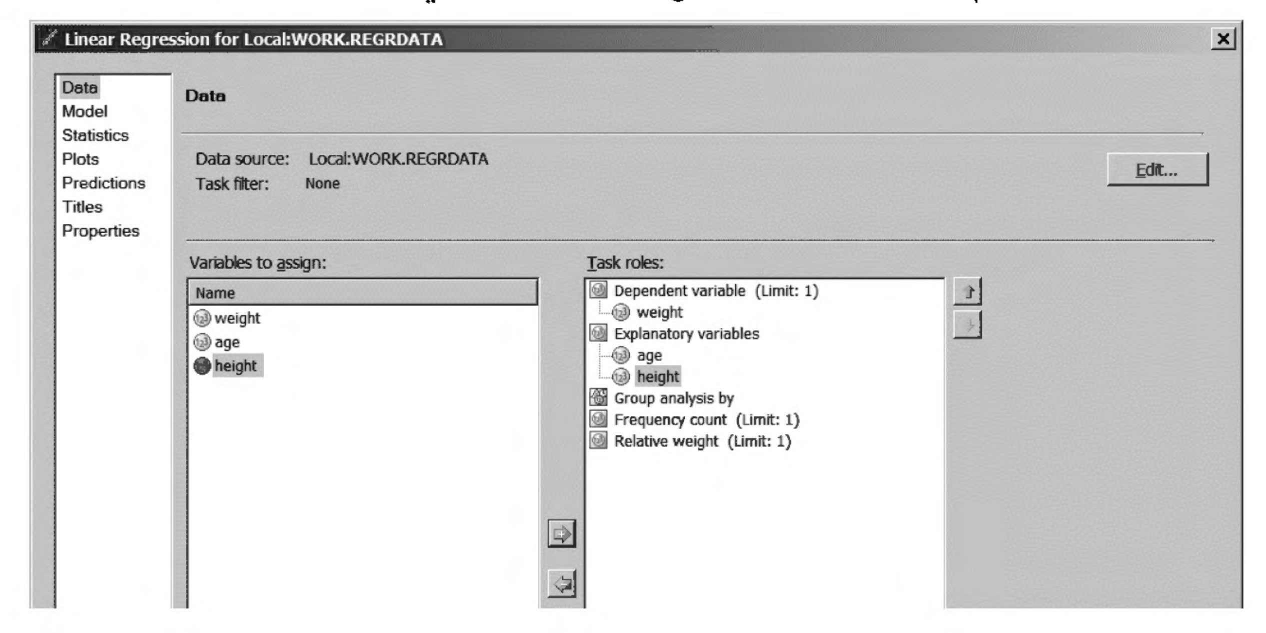


إطار رقم (٣-٦): قائمة Analyze في SAS Enterprise Guide

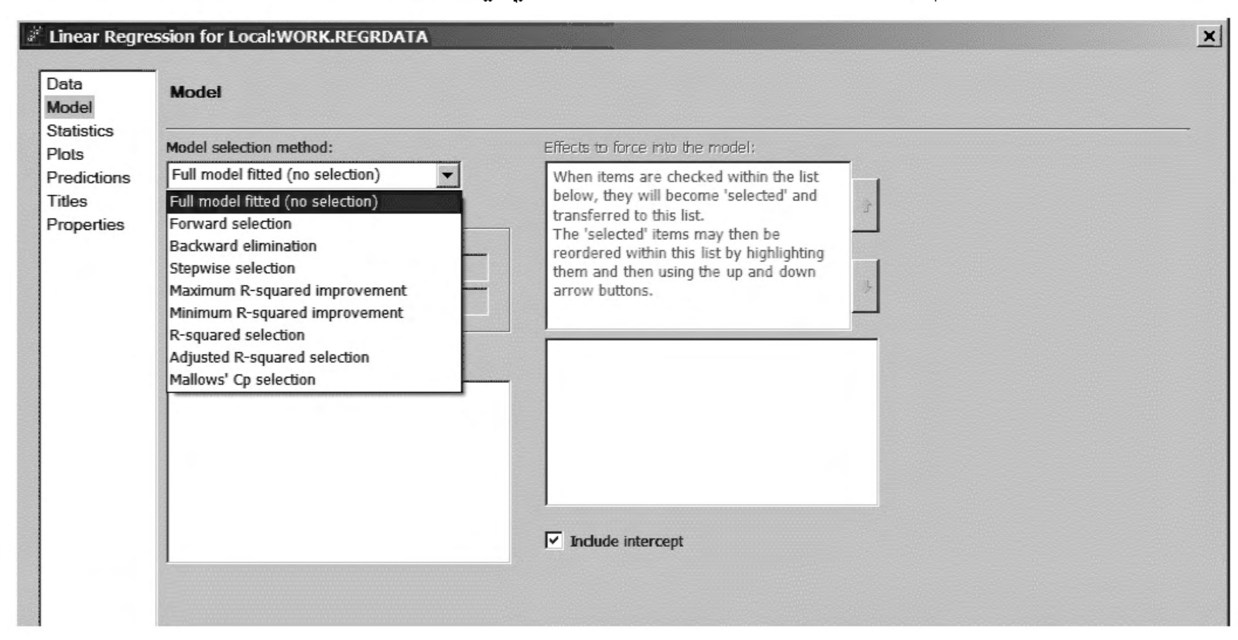


	weight	age	height
1	11.50	3.00	84
2	16.00	5.00	95
3	6.50	0.50	65
4	17.00	4.00	100
5	8.50	1.33	70
6	8.80	1.00	70
7	22.00	6.17	118
8	13.00	3.42	95
9	12.50	3.67	94
10	15.50	5.42	97
11	9.50	1.17	76
12	15.50	4.42	96
13	9.50	1.17	73
14	14.50	2.75	100
15	19.00	6.25	115
16	9.00	1.50	76
17	14.00	4.25	98
18	10.50	2.00	80
19	6.00	0.42	63
20	15.00	5.58	105
21	13.00	3.42	94
22	21.00	6.17	118
23	12.00	3.00	90
24	17.50	5.25	100

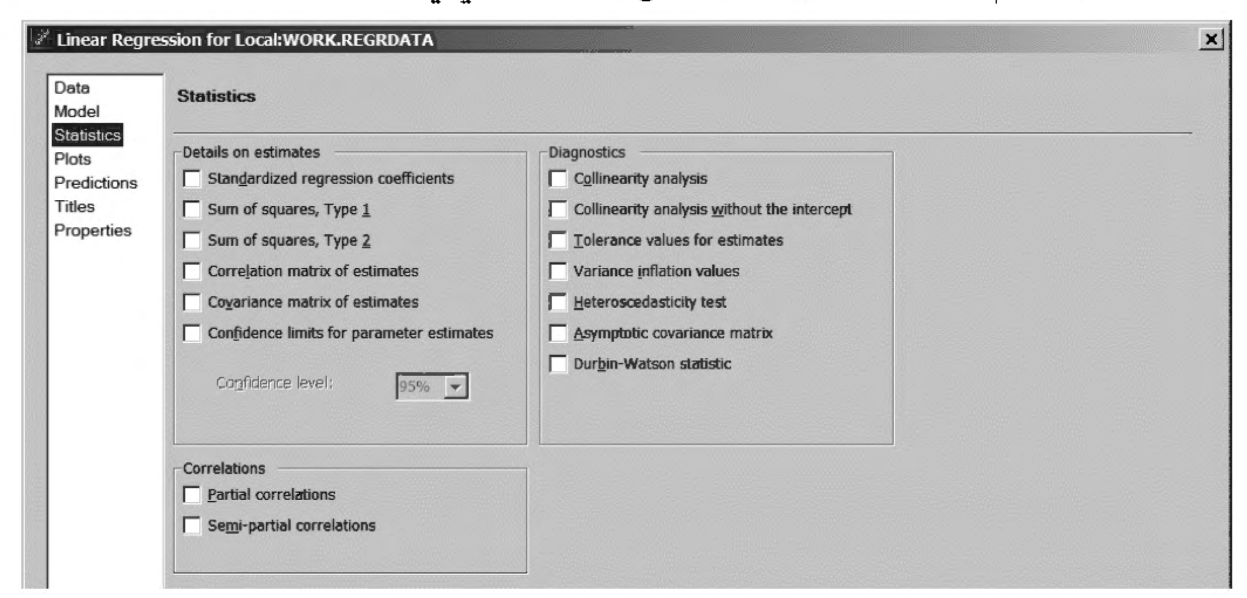
إطار رقم (٣-٧): تحديد المتغير التابع والمتغيرات المستقلة في SAS Enterprise Guide



إطار رقم (٣-٨): خيارات نموذج الانحدار الخطي في SAS Enterprise Guide



إطار رقم (٣-٩): خيارات إحصاءات نموذج الانحدار الخطي في SAS Enterprise Guide



٣-٢٠-١-٣ مثال[‡‡] تحليل انحدار وزن الطفل على العمر والطول باستخدام نظام SAS:

يوضح الإطار رقم (٣-٣) شكل البيانات والأوامر المطلوبة لإجراء تحليل الانحدار باستخدام الإجراء PROC REG. حيث تم أولاً تعريف متغيرات النموذج: الوزن (WEIGHT)، العمر (AGE)، والطول (HEIGHT). وفي الجزء الثاني تمت كتابة الأوامر الخاصة بتحليل الانحدار.

يوضح الإطار رقم (٣-١١) أجزاء المخرجات التي تم الحصول عليها بناء على إجراء تحليل الانحدار الذي تم تحديده في الإطار رقم (٣-٣). حيث تنقسم المخرجات إلى ستة أجزاء كالتالي:

- يتضمن الجزء الأول من المخرجات المجموع (Sum) والوسط الحسابي (Mean) ومجموع المربعات غير المصحح (Uncorrected SS) والتباين (Variance) والانحراف المعياري (Standard Deviation) للمتغيرات الثلاثة (العمر، الطول والوزن) والثابت. حيث توضح النتائج أن متوسط عمر الطفل في عينة الدراسة (٢,٣٥) سنة بانحراف معياري قدره (٢,١) سنة ومتوسط طول الطفل (٧٦,٤) سنتيمتر وانحراف معياري قدره (٢٤,٠١) وأن متوسط الوزن قد بلغ (١٠,١٥٤) كيلوجرام بانحراف معياري (٥,٥٣) كيلوجرام.
- يوضح الجزء الثاني مصفوفة مجموع المربعات غير المصحح بين المتغيرات. وفي الواقع لا توجد فائدة عملية من نتائج هذه المصفوفة.
- أما الجزء الثالث فيوضح مصفوفة معاملات الارتباط الخطي بين المتغيرات الثلاثة. وتشير النتائج إلى وجود علاقة طردية قوية[§§] بين المتغيرات الثلاثة.
- يحتوي الجزء الرابع على جدول تحليل التباين. وتشير نتائج جدول تحليل التباين إلى معنوية الانحدار ككل (Prob>F=0.0001)، أي أن متغيري العمر والطول يسهمان بمستوى معنوي عالٍ من الدلالة في تفسير التغير في وزن الطفل.
- يحتوي الجزء الخامس على قيم خمسة إحصاءات هي: الخطأ المعياري للتقدير (Root MSE) – الانحراف المعياري لقيم المتغير التابع الفعلية عن خط انحدار المقدر المناظر لقيم المتغيرات المستقلة المختلفة-، والوسط الحسابي للمتغير التابع (Dep Mean)، معامل الاختلاف (C.V.)–نسبة الخطأ المعياري للتقدير للوسط الحسابي للمتغير التابع-، ومعامل التحديد (R-square) ومعامل التحديد المعدل (Adj R-sq). وتوضح النتائج أن الخطأ المعياري لخط الانحدار قد بلغ (١,١) مما يشير إلى قرب القيم الحقيقية للمتغير التابع للقيم المقدرة لها، وبلغ

‡‡ مصدر البيانات مستشفى أبها للنساء والولادة والمراكز الصحية (١٩٩٩)؛ بيانات عن أوزان (٥٠) طفلاً (انظر الفصل الثاني).

§§ على الرغم من قوة الارتباط بين متغيري العمر والطول، إلا أنه وفقاً لمعياري عامل تضخم التباين (VIF) الذي بلغ (٧,٩٨٦) ومؤشر الحالة (Condition index) الذي بلغ (١٩,٨٤٨)، فإن النموذج لا يعاني مشكلة الارتباط الخطي المتعدد (Multicollinearity)، وللمزيد حول مشكلة الارتباط الخطي المتعدد يرجى الرجوع إلى الفصل السابع.

الوسط الحسابي لأوزان الأطفال (المتغير التابع) (١٠,١٥٤) كيلوجرام وبلغ معامل الاختلاف (١٠,٨٢%). أما معامل التحديد فيوضح أن (٩٦,٢%) من التغير في وزن الطفل قد جرى تفسيره بواسطة متغيري العمر والطول.

– يحتوي الجزء السادس على مقدرات المعالم (parameter estimate)، الخطأ المعياري للمقدرات (Standard error)، قيم t تحت فرض العدم $\left(H_0: \beta_r = 0,\ r = 0,1,2\right)$، قيم الاحتمال ومجموع المربعات من النوع الثاني. وتوضح النتائج المستعرضة أن كلا من المتغيرين التفسيرين يسهم إسهاما جوهرياً في التنبؤ بوزن الطفل (p-value<<0.01). ويتم عادة تقرير النموذج الموفق*** كما يوضح الإطار رقم (٣-١٠):

إطار رقم (٣-١٠): عرض نتائج نموذج الانحدار الخطي وفق نظام جمعية علم النفس الأمريكية نموذج انحدار وزن الطفل على طوله وعمره (عدد المشاهدات =٥٠)

المتغير	المعامل	الخطأ المعياري	معامل الانحدار المعياري
الطول (سم)	٠,١٢	٠,٠٢	٠,٥٤***
العمر (سنة)	١,٢٠	٠,٢١	٠,٤٦***

معامل التحديد R^2 = ٠,٩٦؛ معامل التحديد المعدل $\bar{R}^2$=٠,٩٦

*** قيمة الاحتمال أقل من ٠,٠٠١ $(p < .001)$

*** انظر الفصل الثامن حول عرض نتائج نموذج الانحدار الخطي.

إطار رقم (٣-١١): مخرجات برنامج SAS لنموذج انحدار وزن الطفل على العمر والطول

The SAS System

The REG Procedure

Number of Observations Read	50
Number of Observations Used	50

الإحصاء الوصفي للمتغيرات

الجزء الأول

Descriptive Statistics					الانحراف المعياري
المتغير Variable	المجموع Sum	الوسط الحسابي Mean	مجموع المربعات غير المصحح Uncorrected SS	التباين Variance	Standard Deviation
Intercept	50.00000	1.00000	50.00000	0	0
age	117.36000	2.34720	491.68580	4.41262	2.10062
height	3820.00000	76.40000	320090	576.36735	24.00765
weight	507.70000	10.15400	6653.38500	30.57549	5.52951

مصفوفة مجموع المربعات غير المصحح بين المتغيرات

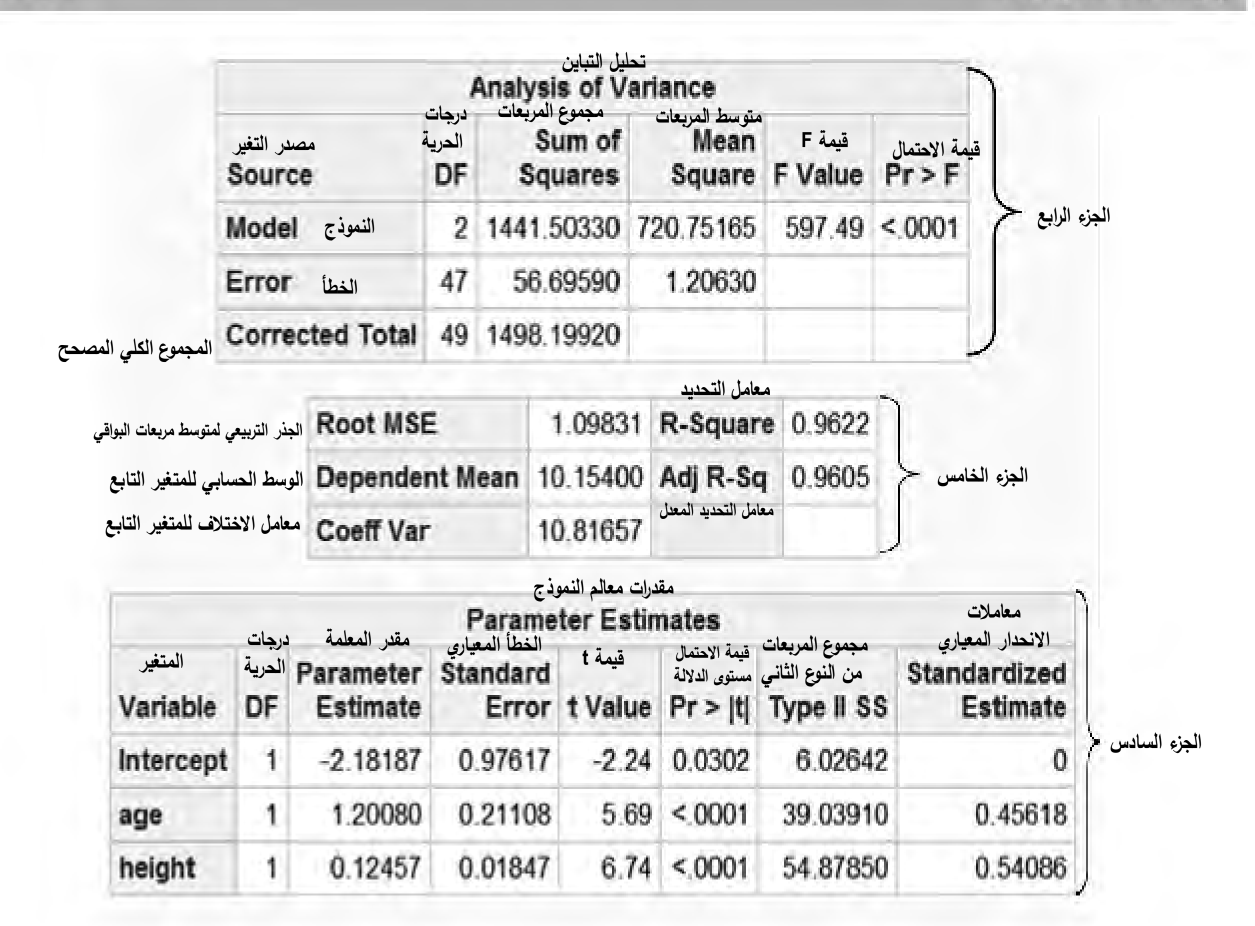
الجزء الثاني

Uncorrected Sums of Squares and Crossproducts				
Variable	Intercept	age	height	weight
Intercept	50	117.36	3820	507.7
age	117.36	491.6858	11277.54	1739.226
height	3820	11277.54	320090	45081.8
weight	507.7	1739.226	45081.8	6653.385

مصفوفة معاملات الارتباط الخطي البسيط بين المتغيرات

الجزء الثالث

Correlation			
Variable	age	height	weight
age	1.0000	0.9353	0.9620
height	0.9353	1.0000	0.9675
weight	0.9620	0.9675	1.0000

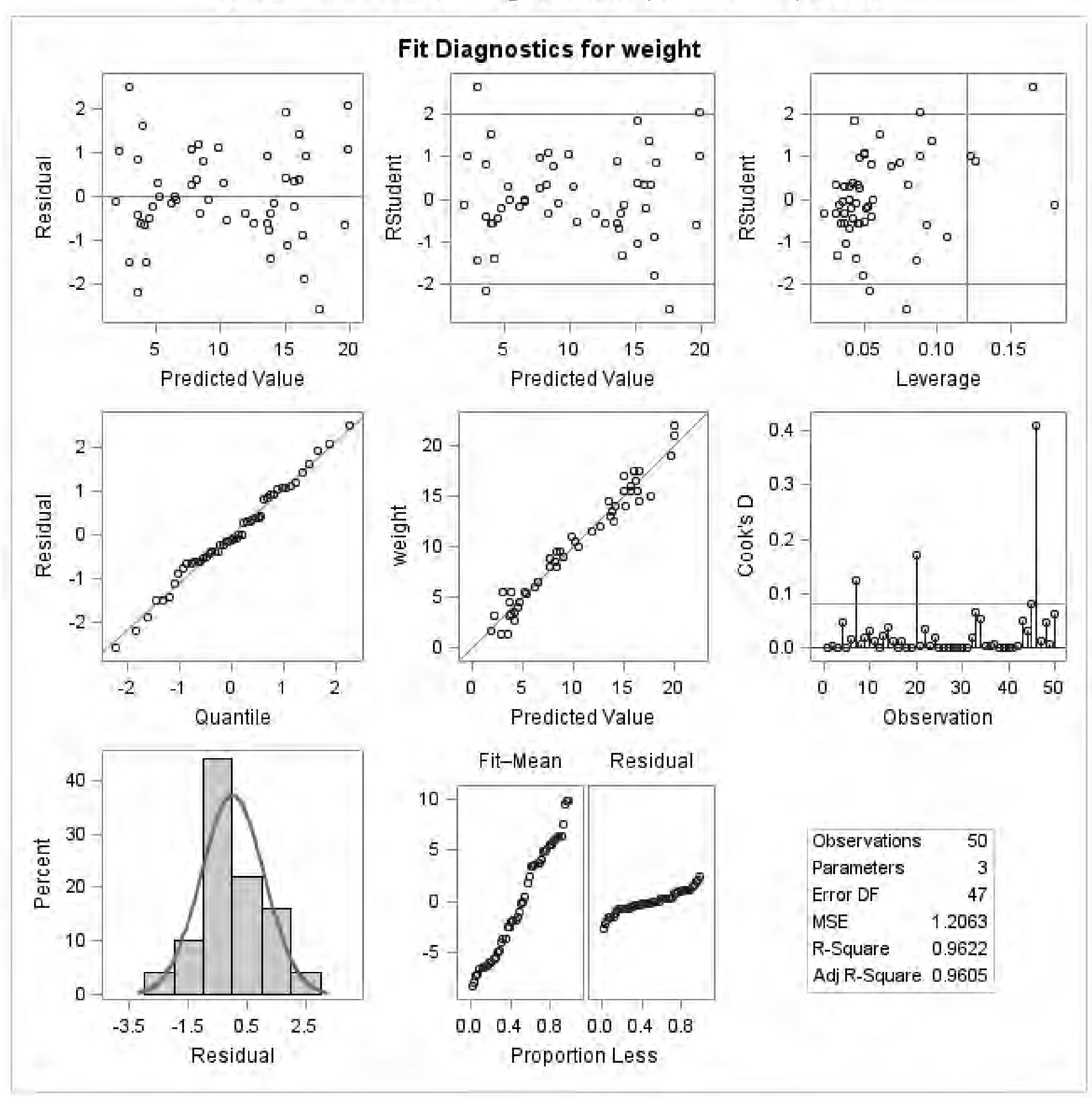


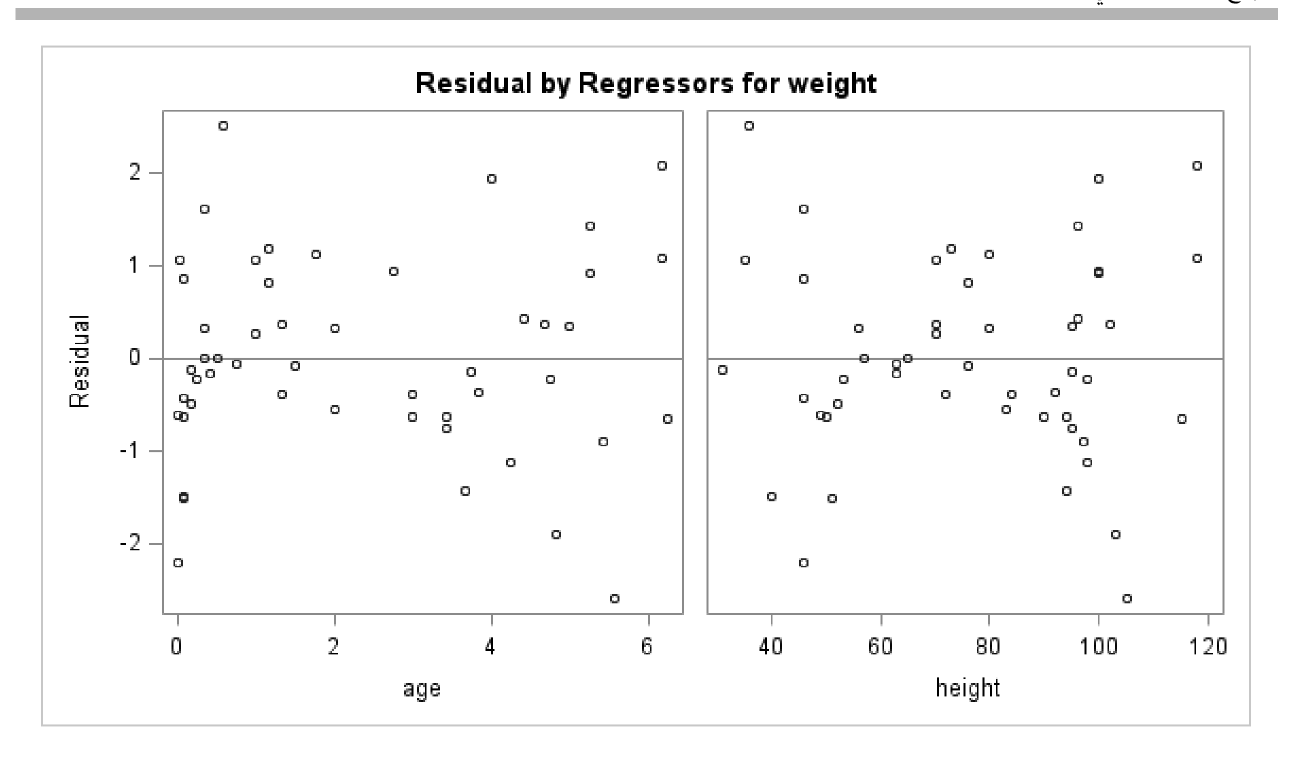
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1441.50330	720.75165	597.49	<.0001
Error	47	56.69590	1.20630		
Corrected Total	49	1498.19920			

Root MSE	1.09831	R-Square	0.9622
Dependent Mean	10.15400	Adj R-Sq	0.9605
Coeff Var	10.81657		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type II SS	Standardized Estimate
Intercept	1	-2.18187	0.97617	-2.24	0.0302	6.02642	0
age	1	1.20080	0.21108	5.69	<.0001	39.03910	0.45618
height	1	0.12457	0.01847	6.74	<.0001	54.87850	0.54086

إطار رقم (٣-١٢): رسوم بيانية للبواقي لفحص مدى ملاءمة النموذج





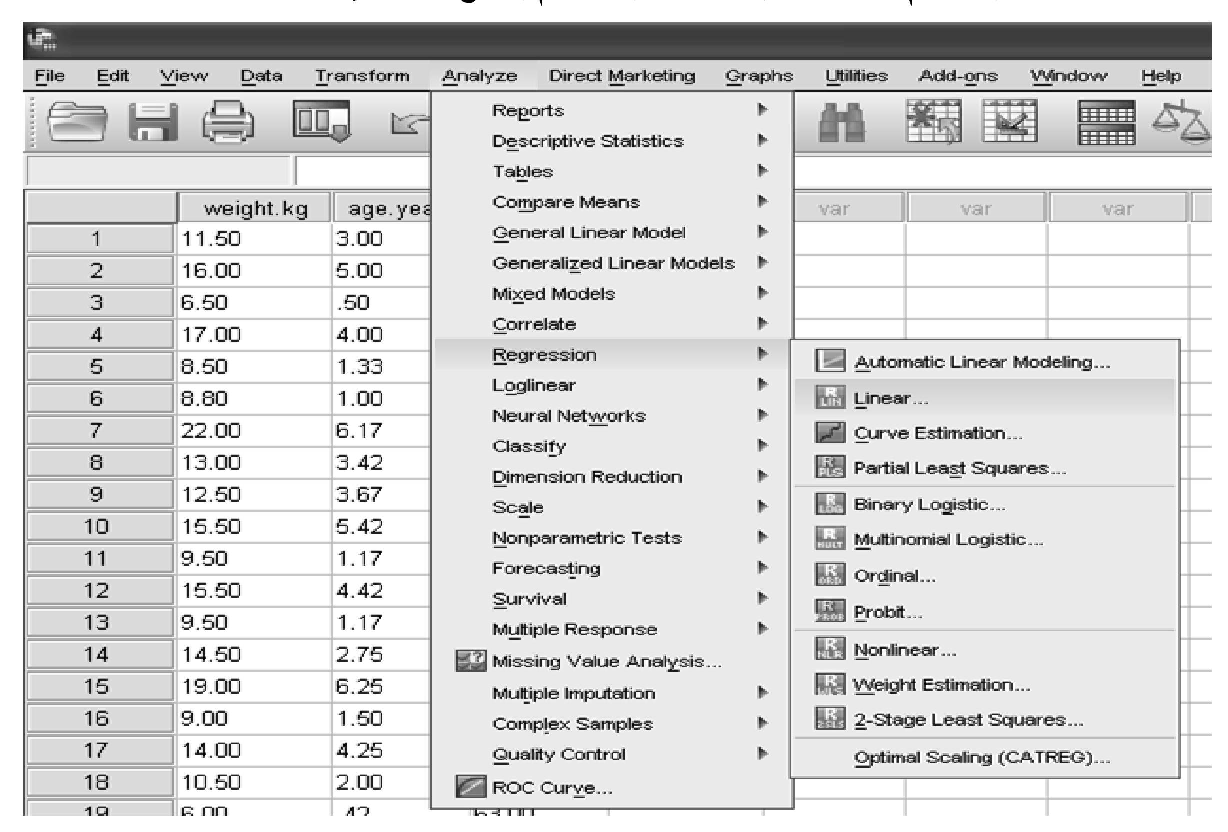
٣-٢٠-٢ برنامج الحزمة الإحصائية للعلوم الاجتماعية (Statistical Package for Social Sciences (SPSS)):

يعد برنامج الحزمة الإحصائية للعلوم الاجتماعية (IBM SPSS) من البرامج الإحصائية الواسعة الاستخدام في مجال التحليل الإحصائي. ويتميز البرنامج SPSS* تحت بيئة النوافذ (Under windows) بسهولة استخدامه الذي لا يتطلب أي معرفة ببرمجة الحاسب الآلي.

ويوضح الإطار رقم (٣-١٣) شكل البيانات والخيارات المطلوبة لإجراء تحليل الانحدار. حيث يتم أولاً إدخال البيانات في جدول إلكتروني أشبه بجدول إكسل يتكون من أعمدة وصفوف، تمثل الأعمدة المتغيرات المراد تحليلها والصفوف الحالات أو المشاهدات. كما يمكن إدخال البيانات في برنامج آخر كبرنامج إكسل ولوتس أو أي برنامج قاعدة بيانات (dBase, Access, etc.) ومن ثم يتم نسخها أو قراءتها مباشرة في برنامج SPSS. ويتم إجراء تحليل الانحدار بعد فتح ملف البيانات باختيار "ANALYZE" من قائمة الخيارات الرئيسية ومن ثم اختيار "REGRESSION" واختيار "LINEAR" ومن بعد ذلك يتم تحديد المتغير التابع والمتغيرات المستقلة في خانتي "Dependent" و"Independents" على التوالي (إطار رقم ٣-١٣).

* للمزيد حول برنامج SPSS يرجى الرجوع إلى باشيوة (٢٠١٣م) ومحمود (٢٠١٣م) و(Field, 2013) و(George & Mallery, 2013)

إطار رقم (٣-١٣) تحليل الانحدار باستخدام برنامج SPSS الإصدار (٢٠)



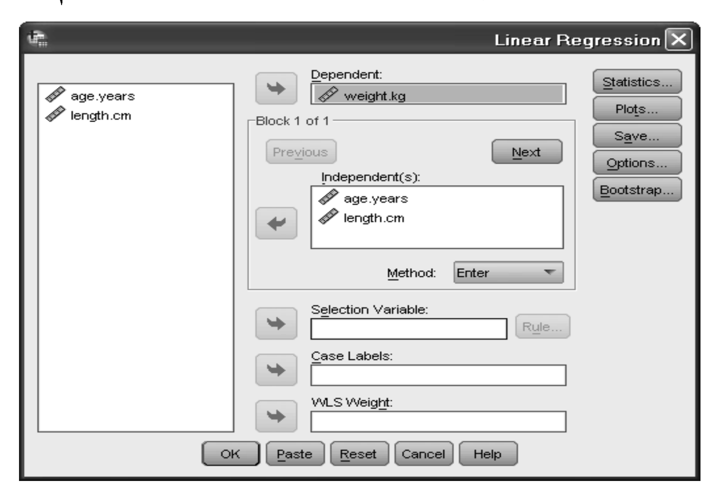
- خيار اختيار المتغيرات المستقلة Method: يتيح هذا الخيار اختيار المتغيرات المستقلة في نموذج الانحدار (إطار رقم ٣-١٤). والخيارات المتاحة هي:
  - Enter: لإدخال المتغيرات المستقلة في النموذج في خطوة واحدة.
  - Remove: لحذف المتغيرات المستقلة من نموذج الانحدار في خطوة واحدة.
  - Stepwise: لإجراء تحليل الانحدار التدرجي.
  - Forward: لإجراء تحليل الانحدار باستخدام طريقة الإضافة إلى الأمام.
  - Backward: لإجراء تحليل الانحدار باستخدام طريقة الحذف إلى الخلف.

- خيار بناء نموذج لحالاتِ محددة وفق متغير تصنيفي Selection variable : يتيح البرنامج بناء نموذج انحدار لحالات محددة بناءً على خاصية محددة مثل نموذج بناء نموذج للإناث فقط وفق متغير الجنس (١=ذكر،٢=أنثي) باختيار متغير الجنس وتحديد القيمة "٢".

- خيار تحديد متغير أسماء الحالات Case label : يتيح البرنامج تحديد متغير لأسماء الحـالات للاسـتفادة منهـا في تحديد أرقام الحالات أو أسمائها في حالة البيانات الشاذة وغيرها.
- WLS: يتيح هذا الخيار إجراء تحليل الانحدار المرجح وذلك بتحديد المتغير الذي يمثل الأوزان "Weights".



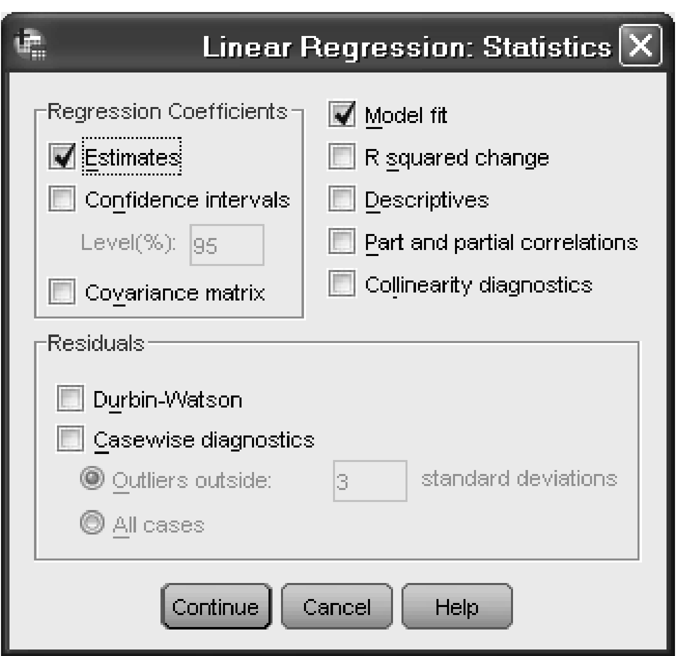
إطار رقم (٣-١٤): تحديد المتغيرات والأوامر لتحليل الانحدار باستخدام برنامج SPSS



- إحصاءات الانحدار Statistics: يتيح هذا الخيار المخرجات التالية (إطار رقم (٣-١٥):
  - مقدرات المربعات الصغرى.
  - فترة ثقة لمقدرات المربعات الصغرى.
  - مصفوفة التباين والتغاير.
  - نتائج النموذج الموفق.
  - التغير في معامل التحديد.
  - الإحصاء الوصفي
  - معامل الارتباط الجزئي.
  - مؤشرات تشخيص مشكلة الارتباط الخطي المتعدد.
  - تحليل البواقي (ديربن - واتسون وتحديد المشاهدات الشاذة).

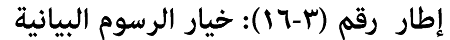


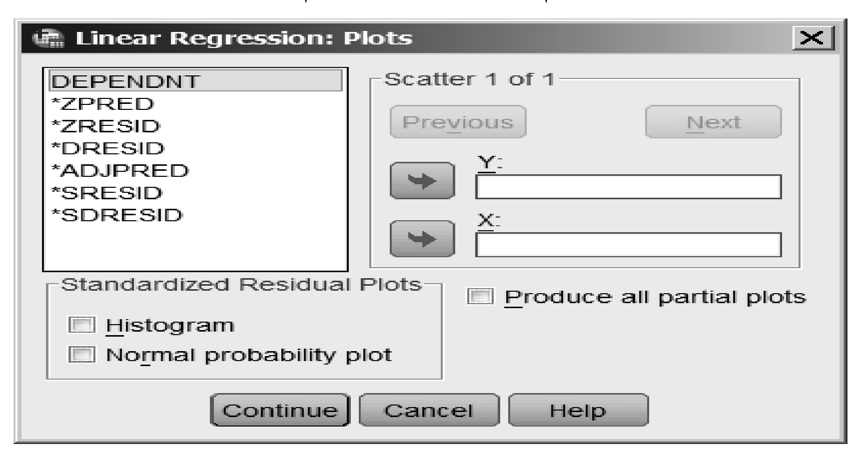
إطار رقم (٣-١٥): تحديد إحصاءات الانحدار



- خيارات الرسوم البيانية (Linear Regression: Plots):

يتيح هذا الخيار (إطار رقم ٣-١٦) رسم مجموعات من الرسوم البيانية التي تستخدم في كأدوات لفحص مدى استيفاء النموذج لإشتراطات الانحدار. فالجزء الأعلى من صندوق الخيارات يتيح رسم مجموعة من الأشكال الانتشارية لأزواج متغيرات المتغير التابع والقيم الموفقة المعيارية ((Standardized predicted values (*ZPRED))، والبواقي المعيارية (Standardized residuals (*ZRESID، والبواقي المحذوفة Deleted residuals (*DRESID), وبواقي القيم الموفقة المعدلة Adjusted predicted values (*ADJPRED)، وبواقي ستيودنت Studentized residuals (*SRESID) وبواقي ستيودنت المحذوفة(Studentized deleted residuals (*SDRESID . وبالإضافة إلى الرسوم الانتشارية، يتيح البرنامج أيضاً رسم المدرج التكراري والاحتمال الطبيعي للبواقي المعيارية للتأكد من مدى تبعية البواقي للتوزيع الطبيعي. كما يتيح البرنامج الرسوم الجزئية (Partial plots).

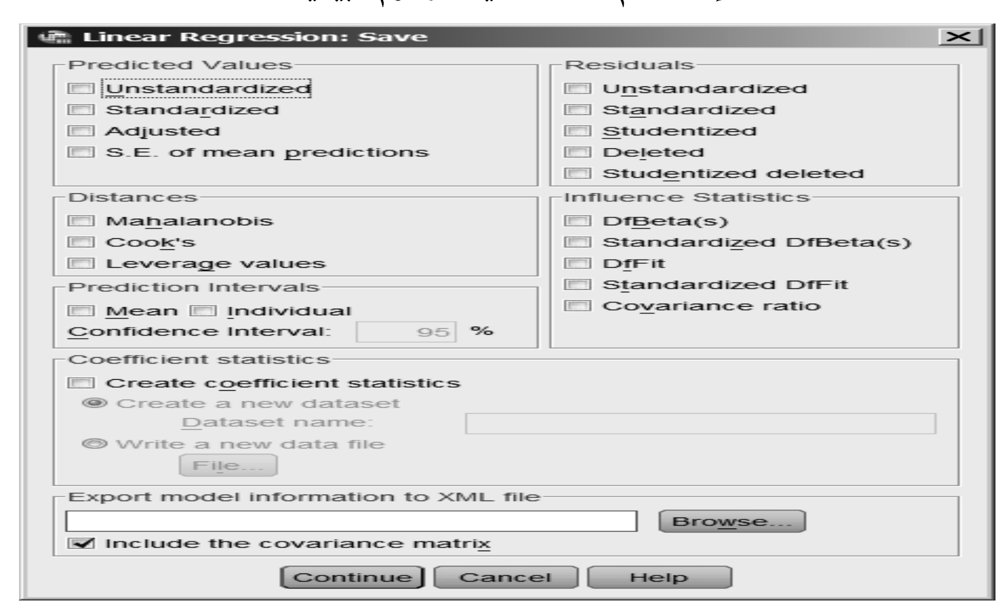

إطار رقم (٣-١٦): خيار الرسوم البيانية



- خيارات الحفظ (Linear Regression: save):

يتيح هذا الخيار - إطار رقم (٣-١٧)- حفظ القيم الموفقة (Predicted values)، والبواقي (Residuals)، مقاييس المسافة (Distances) التي تستخدم لتحديد القيم الشاذة، وإحصاءات التأثير (Influence statistics) وإحصاءات معاملات الانحدار (Coefficient statistics)، تصدير بيانات النموذج لملف XML ومصفوفة التباين والتغاير (Covariance matrix).

إطار رقم (٣-١٧): خيار الرسوم البيانية



- خيارات الانحدار الخطي (Linear Regression Options):

-معايير طرق اختيار المتغيرات (Stepping Method Criteria): يتيح النظام خيار تحديد مستوى المعنوية لدخول المتغيرات المستقلة في النموذج بالنسبة لطريقتي الإضافة إلى الأمام والانحدار التدرجي وكذلك تحديد مستوي المعنوية لخروج المتغير المستقل من النموذج في طريقة الحذف من الخلف؛ أو خيار تحديد قيمة F لدخول وحذف المتغيرات المستقلة. كما يلاحظ أن قيم مستوى المعنوية المحددة بواسطة النظام هي (٠,٠٥) لطريقتي الإضافة إلى الأمام والانحدار التدرجي و(٠,١٠) لطريقة الحذف من الخلف وفي حالة اختيار قيم F كمعيار للاختيار نجد أن القيم المحددة بواسطة النظام هي (٣,٨٤) و(٢,٧١) لإدخال وخروج المتغيرات المستقلة على التوالي (إطار رقم ٣-١٨).

- خيار إدخال المعامل الثابت في النموذج (Include constant in equation).
- القيم المفقودة (Missing values): يتيح النظام ثلاثة خيارات لمعالجة القيم المفقودة، هي: استبعاد الحالات من القائمة Exclude cases listwise، استبعاد الحالات زوجياً Exclude cases pairwise، استبدال القيم المفقودة بالوسط الحسابي لبقية القيم Replace with mean.

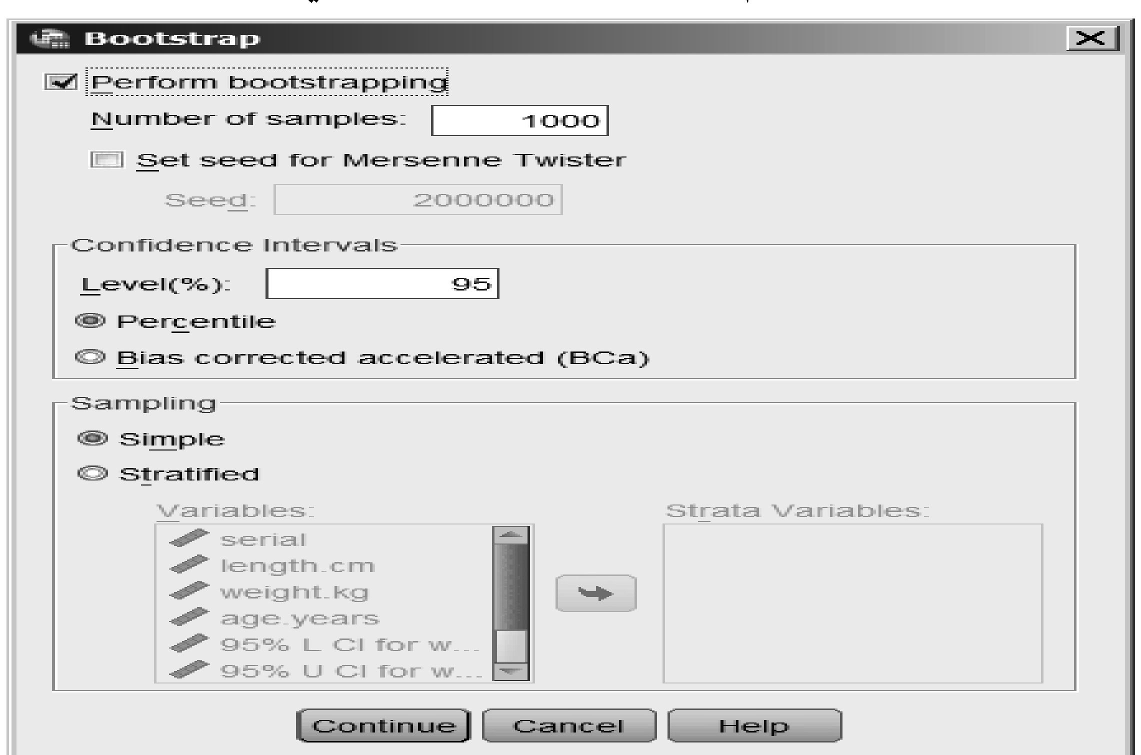
إطار رقم (٣-١٨): خيارات نموذج الانحدار الخطي

Linear Regression: Options

Stepping Method Criteria
Use probability of F
Entry: .05 Removal: .10
Use F value
Entry: 3.84 Removal: 2.71
Include constant in equation
Missing Values
Exclude cases listwise
Exclude cases pairwise
Replace with mean
Continue Cancel Help

- خيار البوتستراب[†††] (Bootstrap): يعد هذا الخيار (إطار رقم ٣-١٩) من أهم الخيارات المضافة في الإصدارات الحديثة لبرنامج SPSS. وتتيح طريقة البوتستراب الحصول على مقدرات حصينة (robust) للأخطاء المعيارية وفترات الثقة لمعاملات الانحدار، لذا يفضل استخدامها في حالة صغر حجم العينة وفي وجود مشكلة عدم ثبات التباين (Heteroscedasticity).

إطار رقم (٣-١٩): خيارات نموذج الانحدار الخطي



مثال: تحليل انحدار وزن الطفل على العمر والوزن باستخدام نظام SPSS

وضح الإطار (٣-٢٠) مخرجات برنامج SPSS لنموذج انحدار وزن الطفل (Weight) على العمر (Age) والطول (Height). وتم تقسيم مخرجات النموذج إلى أربعة أجزاء، هي:

الجزء الأول: تعتمد مخرجات هذا الجزء على طريقة اختيار المتغيرات (Method) المختارة (انظر الإطار ٣-١٧). ففي هذا المثال تم اختيار طريقة الاختيار "ENTER" لإدخال المتغيرين -العمر والطول- في النموذج.

[†††] للمزيد حول تقديرات البوتستراب يرجي الرجوع Montgomery, Peck and Vining (2012) .

الجـــزء الثـــاني: تعتمـــد مخرجـــات هـــذا الجـــزء عـــلى اختيـــار (Model fit) مـــن إحصـــاءات الانحـــدار "Linear Regression: Statistics" (إطار (٣-١٥)). وإحصاءات هذا الجزء هي:

- معامل الارتباط المتعدد(R) : ويشير الارتباط المتعدد البالغ قدره ٠,٩٨١ إلى أن العلاقة بين القيم المقدرة والقـيم الفعلية قوية جداً مما يدل إلى حسن جودة التوفيق.
- معامل التحديد (R square): يوضح معامل التحديد أن متغيري العمر والطول قد فســرا ٩٦,٢% مـن التغـير في الوزن.
- معامل التحديد المعدل (Adjusted R square): بلغ معامل التحديد المعدل ٩٦,١%.
- الخطأ المعياري للتقدير (Std. Error of the estimate): بلغ الخطأ المعياري للتقدير ١,٠٩٨٣.

الجزء الثالث: في هذا الجزء تم حساب جدول تحليل التباين (ANAOVA). حيث يعـرض العمـود الأول مـن الجـدول مصدر التغير (الانحدار، البـواقي والمجمـوع)، وفي العمـود الثـاني مجمـوع المربعـات، الثالـث درجـات الحريـة، والرابـع متوسط المربعات والخامس قيمة إحصاء "t" ويحتوي العمود الأخير على مستوى المعنوية.

وبما أن مستوي المعنوية أقل بكثير من ٠,٠٠١، فإننا نرفض فرض العدم $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$ عند مستوى عالي من الدلالة، بمعنى أن متغيري الطول والعمر يؤثران على وزن الطفل.

الجزء الرابـع: يعـرض هـذا الجـزء مقـدرات المربعـات الصـغرى لمعـالم النمـوذج. حيـث يحتـوي العمـود الأول مـن المخرجات على أسماء المتغيرات بما في ذلك المعامل الثابت، والعمـود الثـاني عـلى قـيم معـاملات النمـوذج المقـدرة غـير المعيارية والعمود الثالث على الأخطاء المعيارية المناظرة للمعاملات المقدرة، والعمـود الرابـع عـلى معـاملات النمـوذج المعيارية (Standardized coefficients)، والعمود الخامس على قيم t المنـاظرة لمعـاملات النمـوذج المقـدرة تحـت فـرض العدم، أي

$$t_j = \frac{\hat{\beta}_j - 0}{s.e.(\hat{\beta}_j)} \qquad \text{for } j = 0,1,2$$

ويوضح العمود الأخير على مستوى المعنوية. وبما أن قيم مستوى المعنوية للمتغيرين أقل بكثير من ٠,٠٠١، نـرفض كلاً من الفرضين الصفرين ($H_0{:}\beta_2 = 0,\ \ H_0{:}\beta_1 = 0$) عند مستوى عال من الدلالـة. وهـذا يعنـي أن كـلا المتغـيرين التفسيرين - العمر والطول - يسهم إسهاماً جوهرياً في التنبؤ بوزن الطفل.

إطار رقم (٣-٢٠): مخرجات برنامج SPSS لنموذج انحدار وزن الطفل على العمر والطول.

المتغيرات المدخلة / المستبعدة

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	العمر, الطول[b]	.	Enter

a. Dependent Variable: الوزن

b. All requested variables entered.

موجز نتائج النموذج

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.981[a]	.962	.961	1.09831

a. Predictors: (Constant), العمر, الطول

جدول تحليل التباين

ANOVA[a]

Model	النموذج	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1441.503	2	720.752	597.492	.000[b]
	Residual	56.696	47	1.206		
	Total	1498.199	49			

a. Dependent Variable: الوزن

b. Predictors: (Constant), العمر, الطول

معاملات النموذج

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.182-	.976		-2.235-	.030
	الطول	.125	.018	.541	6.745	.000
	العمر	1.201	.211	.456	5.689	.000

a. Dependent Variable: الوزن

٣-٢٠-٣ برنامج إكسل (EXCEL):

تُعَدّ برامج الجداول الإلكترونية من أوسع برامج الحاسبات الشخصية انتشاراً. ويعد برنامج إكسل من أفضل هذه البرامج بما له من إمكانات كبيرة واستخدامات متنوعة. ويستخدم برنامج إكسل لإجراء العمليات الحسابية الرياضية المختلفة. كما يتيح البرنامج تحليل البيانات الإحصائية من خلال مجموعة من أدوات تحليل البيانات تسمى بـ

"Analysis Toolpack". ومن بين هذه الأدوات الإحصائية التي يتيحها البرنامج أداة تحليل الانحدار الخطي باستخدام طريقة المربعات الصغرى. ولإجراء تحليل الانحدار الخطي يتم اختيار "Data Analysis" من قائمة "أدوات-Tools" ومن ثم يتم اختيار "Regression"[‡‡‡]. ولإجراء نموذج الانحدار يتم تحديد نطاق المتغير التابع ونطاق المتغير/المتغيرات المستقلة في حدود (٦٤) متغيراً. ويتيح البرنامج مجموعة من الخيارات (إطار رقم (٣-٢١)):

- Label: لتحديد نطاق أسماء المتغيرات.
- Constant is Zero: خيار عدم إدراج المعامل الثابت في النموذج.
- Confidence level: خيار حساب فترة الثقة لمعالم نموذج الانحدار المقدرة.
- Residuals: يتيح هذا الخيار مجموعة من المخرجات تشمل: القيم الموفقة، البواقي، البواقي المعيارية ورسوم انتشار البواقي مع كل متغير من المتغيرات المستقلة، رسم انتشار القيم الموفقة مع القيم الفعلية ورسم الاحتمال الطبيعي للبواقي.

يعرض الإطار (٣-٢٢) مخرجات برنامج EXCEL لنموذج انحدار وزن الطفل (Weight) على العمر (Age) والطول (Height). ويتكون مخرجات النموذج إلى ثلاثة أجزاء، هي:

إحصاءات الانحدار (Regression Statistics): يحتوي هذا الجزء على إحصاءات معامل الارتباط المتعدد، معامل التحديد، معامل التحديد المعدل، الخطأ المعياري للتقدير وعدد المشاهدات.

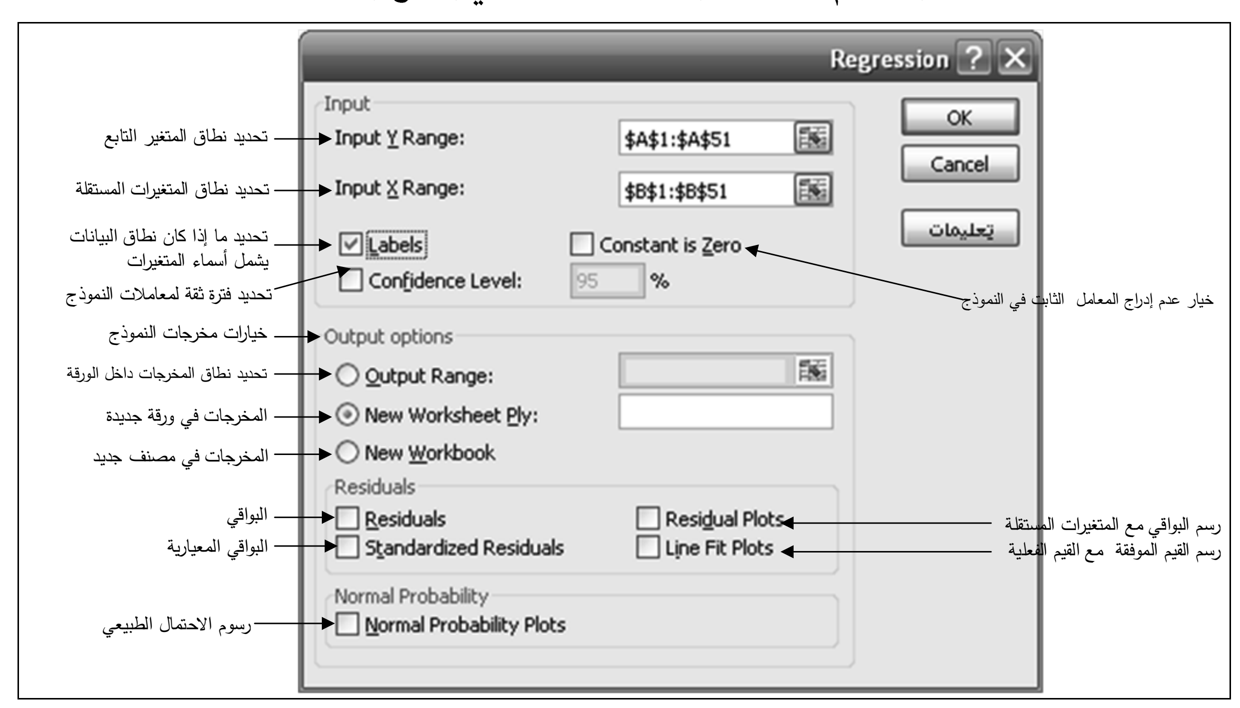
جدول تحليل التباين (ANOVA): في هذا الجزء تم حساب جدول تحليل التباين، حيث يعرض العمود الأول من الجدول مصدر التغير (الانحدار، البواقي والمجموع)، وفي العمود الثاني درجات الحرية، الثالث مجموع المربعات، والرابع متوسط المربعات والخامس قيمة إحصاء "t" ويحتوي العمود الأخير على مستوى المعنوية.

معاملات الانحدار المقدرة: يعرض هذا الجزء مقدرات المربعات الصغرى لمعالم النموذج. حيث يحتوي العمود الأول من المخرجات على أسماء المتغيرات بما في ذلك المعامل الثابت، والعمود الثاني على قيم معاملات النموذج المقدرة، والعمود الثالث على الأخطاء المعيارية المناظرة للمعاملات المقدرة، والعمود الرابع على قيم إحصاء "t" المناظرة لمعاملات النموذج المقدرة تحت فرض العدم.

---

‡‡‡ كما يمكن استخدام الدوال linest، trend، forecast لبناء نموذج الانحدار الخطي البسيط والمتعدد (للمزيد حول هذه الدوال انظر (2012) Carlberg).

إطار رقم (٣-٢١): خيارات أداة الانحدار في برنامج إكسل



إطار رقم (٣-٢٢): مخرجات برنامج إكسل لنموذج انحدار وزن الطفل على العمر والطول

Regression Statistics	
Multiple R	0.98090
R Square	0.96216
Adjusted R Square	0.96055
Standard Error	1.09831
Observations	50

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1441.50330	720.75165	597.49163	0.00000
Residual	47	56.69590	1.20630		
Total	49	1498.19920			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2.18187	0.97617	-2.23513	0.03020	-4.14567	-0.21807
العمر	1.20080	0.21108	5.68883	0.00000	0.77616	1.62544
الطول	0.12457	0.01847	6.74488	0.00000	0.08742	0.16173

تمارين

١) يعطي الجدول التالي بيانات افتراضية عن الأداء الوظيفي وبعض المتغيرات المؤثرة عليه لعينة عشوائية قوامها (٣٣) موظفاً من منسوبي شركة ما.

م	الأداء الوظيفي (من ١٠٠ درجة)	عدد سنوات التعليم	خبرة الموظف (سنة)	مرتبة الموظف
	Y	X_1	X_2	X_3
1	95	19	19	14
2	75	12	12	7
3	86	14	15	9
4	45	5	4	4
5	65	9	6	5
6	56	8	6	4
7	78	11	15	10
8	56	8	7	4
9	67	10	9	6
10	66	9	7	5
11	68	10	10	6
12	66	9	8	6
13	72	11	11	7
14	87	12	17	10
15	90	17	16	11
16	88	15	18	9
17	78	13	14	11
18	88	16	17	11
19	89	16	15	10
20	64	7	5	5
21	84	13	15	8
22	67	10	10	6
23	69	10	10	8
24	72	12	12	7
25	75	12	13	8
26	75	12	14	12
27	90	16	14	11
28	86	14	16	10
29	93	17	20	13
30	94	18	21	14
31	91	16	17	12
32	51	6	4	4
33	52	7	6	4

ومن بيانات الجدول تم الحصول على المصفوفتين التاليتين:

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{pmatrix} 0.436237 & -0.062870 & 0.028290 & -0.000095 \\ -0.062870 & 0.021228 & -0.010251 & -0.007963 \\ 0.028290 & -0.010251 & 0.014908 & -0.010710 \\ -0.000095 & -0.007963 & -0.010710 & 0.027515 \end{pmatrix}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{pmatrix} 2478 \\ 31150 \\ 32316 \\ 21599 \end{pmatrix}$$

المطلوب:

- باستخدام الصيغة (3.17) تحصل على قيم مقدرات المربعات الصغرى العادية باستعمال النموذج التالي:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\varepsilon$$

- اختبر فرض أن أي من معاملات الانحدار يساوي الصفر عند مستوى دلالة/ معنوية ١%.

- أوجد فترة ثقة (٩٥%) لكل معلمة من معالم نموذج الانحدار.

- احسب التنبؤ بالقيمة المتوسطة للأداء الوظيفي المقابل للقيم التالية:

$$X_1=10, \qquad X_2=6, \qquad X_3=9$$

واحسب فترة التنبؤ بالقيمة المتوسطة عند مستوى دلالة/ معنوية ٥%.

٢) من السؤال الأول احسب معاملات الانحدار المعياري باستخدام طريقة الانحراف المعياري وباستخدام طريقة المدى الربيعي وقارن بين النتيجتين؟

٣) من السؤال الأول، استخدم مبدأ مجموع المربعات الإضافي لاختبار معنوية كل معامل من معاملات الانحدار عند مستوى معنوية؟

٤) من السؤال الأول احسب معاملات الارتباط بين المتغيرات المستقلة والمتغيرات التابع، ومن ثم احسب معاملات الارتباط الجزئي والتحديد الجزئي التالية:

- الأداء الوظيفي ومستوى التعليم.
- الأداء الوظيفي ومدة خبرة الموظف.
- الأداء الوظيفي والمرتبة.

٥) استخدم طرق تحليل البواقي بما في ذلك رسوم الانحدار الجزئية للكشف عن وجود أي انحرافات عن فروض نموذج الانحدار اللازم توافرها؟

٦) وضح لماذا تزيد قيمة معامل التحديد (R^2) كلما تمت إضافة متغير لمعادلة الانحدار؟

٧) إن معامل الارتباط الجزئي بين X_1 وY بعد استبعاد أثر المتغيرات المستقلة ($X_2, X_3, ..., X_p$) هو معامل الارتباط الخطي البسيط بين بواقي نموذج انحدار Y على المتغيرات المستقلة باستثناء المتغير X_1 مع بواقي نموذج انحدار X_1 على بقية المتغيرات المستقلة ($X_2, X_3, ..., X_p$). باستخدام هذه الطريقة، احسب معاملات الارتباط الجزئي والتحديد الجزئي التالية:

- الأداء الوظيفي ومستوى التعليم.
- الأداء الوظيفي ومرتبة الموظف.

الفصل الرابع

المشاهدات الشاذة في تحليل الانحدار الخطي:
طرق كشفها وقياس تأثيرها ومعالجتها

٤-١ مقدمة:

نواجه أحياناً في تحليل الانحدار وجود عدد قليل من المشاهدات الشاذة/ الخارجة والمتطرفة (Outlying & extreme observations) سواء في المتغيرات التابعة أو المتغيرات المستقلة أو المتغيرات التابعة والمستقلة معًا. والمشاهدات الشاذة كما سبق تعريفها هي مجموعة قليلة من المشاهدات تبعد قيمها بصورة كبيرة عن بقية قيم المشاهدات في العينة. ويرجع بروز البيانات الشاذة إلى أخطاء إما في مرحلة جمع البيانات أو في مرحلة المعالجة كإدخال البيانات في الحاسب، وقد تكون هذه البيانات حقيقية ناتجة عن ظروف غير عادية؛ فمثلاً حدوث كوارث طبيعية كالزلازل، والأعاصير، والأمطار الغزيرة يؤثر على مستويات الإنتاج الزراعي، الحيواني والصناعي، إضراب عمال في منشأة ما يؤثر على إنتاجها، الحروب بين الدول تؤثر على اقتصاديات هذه الدول.. وهكذا تكون آثار الكوارث على اختلافها.

إن وجود قيم شاذة/خارجة في مشاهدات المتغير التابع أو المتغيرات المستقلة يؤثر على تقديرات معالم نموذج الانحدار والإحصاءات المرتبطة بها ويؤثر أيضاً على اختيار المتغيرات في نموذج الانحدار (Kleinbaum et al (1988) p. 197). ولقد تطرقنا في الفصل الثالث بإيجاز إلى كيفية استخدام أشكال انتشار البواقي المعيارية مع كل من المتغير التابع والمتغيرات المستقلة في الكشف عن وجود مشاهدات شاذة. وفي هذا الفصل سنتناول بشيء من التفصيل استخدام بعض الطرق التحليلية لمعالجة موضوع الحالات الشاذة من حيث تحديدها وقياس أثرها على مقدرات المربعات الصغرى وطرق معالجتها.

٤-٢ طرق كشف المشاهدات الشاذة:

٤-٢-١ المشاهدات الشاذة في المتغيرات المستقلة:

الرافعة Leverage:

لتحديد الحالات الشاذة في المتغيرات المستقلة، تستخدم عادة مصفوفة القبعة (Hat Matrix) حيث تُعَدُّ العناصر القطرية (h_{ii}) مؤشرًا مفيدًا في الكشف عن الحالات الشاذة. ويسمى العنصر القطري (h_{ii}) بقيمة القبعة (Hat value) ويعرف أيضاً بالرافعة (Leverage). وتقيس الرافعة المسافة بين الحالة رقم i ($x_{1i}, x_{2i},\ldots,x_{pi}$) ومتوسط قيم كل الحالات أو المركز Centroid $(\bar{x}_1, \bar{x}_2,\ldots, \bar{x}_p)$.

ويتم الحصول على قيم القبعة كما يلي:

$$h_{ii} = \mathbf{x}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{x}_i \qquad (4\text{-}1)$$

حيث إن $\mathbf{x}_i^T$ الصف رقم i من المصفوفة $\mathbf{X}$.

وتُراوح قيمة h_{ii} بين $1/n$ وواحد صحيح (Belsley, Kuh and Welsh (1980) pp. 66-67)، أي:

$$\frac{1}{n} \le h_{ii} \le 1 \qquad (4\text{-}2)$$

ومن المعادلة (4.2) نجد أن قيمة القبعة تُراوح بين الواحد الصحيح وحجم العينة $\left(1 \le h_{ii} \le n\right)$. ومجموع قيم العناصر القطرية لمصفوفة القبعة $\mathbf{H} = \mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}$ يساوي عدد معالم نموذج الانحدار بما في ذلك المعامل الثابت، أي أن:

$$\sum_{i=1}^{n} h_{ii} = \operatorname{trace}(\mathbf{H}) = \operatorname{trace}\left(\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\right) = p+1 \qquad (4\text{-}3)$$

حيث p عدد المتغيرات المستقلة في نموذج الانحدار وn عدد المشاهدات.

ومن ثم فإن متوسط قيم القبعة يساوى عدد معالم نموذج الانحدار على عدد المشاهدات، أي أن:

$$\bar{h} = \frac{p+1}{n}$$

وتُعَدُّ قيمة الرافعة كبيرة إذا كانت أكبر من ضعف متوسط قيم الرافعات (Belsley et al (1980) p.17) أي عندما تكون:

$$h_{ii} > \frac{2(p+1)}{n} \qquad (4\text{-}4)$$

ملاحظات:

- تُعَدُّ الرافعة التي تزيد قيمتها على (0.5) كبيرة، وتشير إلى أن الحالة شاذة وتستدعي دراساتها بغض النظر عن عدد المتغيرات المستقلة وحجم العينة ((Neter et al (1990) p.396; Huber (1981)).
- في حالة العينات الصغيرة يتوقع ترشيح عدد كبير من الحالات الشاذة لفحصها. ولذلك يقترح جون فوكس (Fox (1997) p. 280) $\left(h_{ii} > \frac{3(p+1)}{n}\right)$ بدلاً من دراسة الحالات التي تزيد قيم رافعاتها (h_{ii}) عن ضعف متوسط قيم الرافعات $\left(h_{ii} > \frac{2(p+1)}{n}\right)$.
- تكون قيمة القبعة أو الرافعة مساوية للواحد الصحيح ($h_{ii}=1$) عندما تكون قيمة المتغير التابع (y_i) مساوية للقيمة الموفقة $\left(\hat{y}_i\right)$ أي عندما تكون قيمة الباقي تساوي صفراً ($e_i = y_i - \hat{y}_i = 0$).

مسافة مهالانوبس Mahalanobis' distance:

تقيس مسافة مهالانوبس بُعد مشاهدات المتغيرات المستقلة عن قيم الوسط الحسابي المقابلة لها، أي البعد من مركز القيم. وتستخدم مسافة مهالانوبس للكشف عن المشاهدات الشاذة في المتغيرات المستقلة. وتأخذ المسافة الصيغة التالية:

$$MD_i = \left(\mathbf{x}_i - \bar{\mathbf{x}}\right)^{T}\mathbf{S}^{-1}\left(\mathbf{x}_i - \bar{\mathbf{x}}\right) \qquad (4\text{-}5)$$

حيث إن:

$\mathbf{x}_i$ متجه قيم المتغيرات المستقلة للمشاهدة رقم i،

$\mathbf{\bar{x}}$ متجه قيم الوسط الحسابي للمتغيرات المستقلة،

$\mathbf{S}^{-1}$ معكوس مصفوفة التباين والتغاير لقيم المتغيرات المستقلة.

ومن المعادلة (4.5) يمكن اشتقاق **مسافة مهالانوبس** في حالة متغير مستقل واحد وفقاً الصيغة التالية:

$$\mathrm{MD_i} = \left(\frac{x_i - \bar{x}}{S} \right)^2 \qquad (4\text{-}6)$$

حيث إن x_i قيمة المشاهدة رقم i، و $\bar{x}$ قيمة الوسط الحسابي وS الانحراف المعياري للمتغير المستقل.

كما يمكن الحصول على قيم **مسافة مهالانوبس** بدلالة قيمة القبعة (h_{ii}) من المعادلة التالية :(Raykov and Marcoulides, 2008, p.76)

$$\mathrm{MD_i} = (n-1)\left(h_{ii} - \frac{1}{n} \right) \qquad (4\text{-}7)$$

ويلاحظ أن قيمة **مسافة مهالانوبس** تتناسب طردياً مع قيمة الرافعة. وتشير القيم الكبيرة **لمسافة مهالانوبس** إلى أن الحالة شاذة قد تؤثر في مقدرات معالم النموذج والإحصاءات المرتبطة بها. وحيث إن القيمة $(\mathbf{x}_i - \mathbf{\bar{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{\bar{x}})$ تتبع توزيع مربع كاي بدرجات حرية مساوية لعدد المتغيرات المستقلة $(\mathrm{MD_i} \sim \chi^2_{\alpha,p})$، يتم مقارنة قيمة مسافة مهالانوبس بالقيمة الحرجة لتوزيع مربع كاي بدرجات الحرية p ومستوى دلالة محدد.

مثال:

من مثال نموذج انحدار وزن الطفل على متغيري العمر والطول حدد المشاهدات الشاذة في المتغيرين المستقلين باستخدام قيم القبعة/الرافعة ومسافة مهالانوبس؟

الحل:

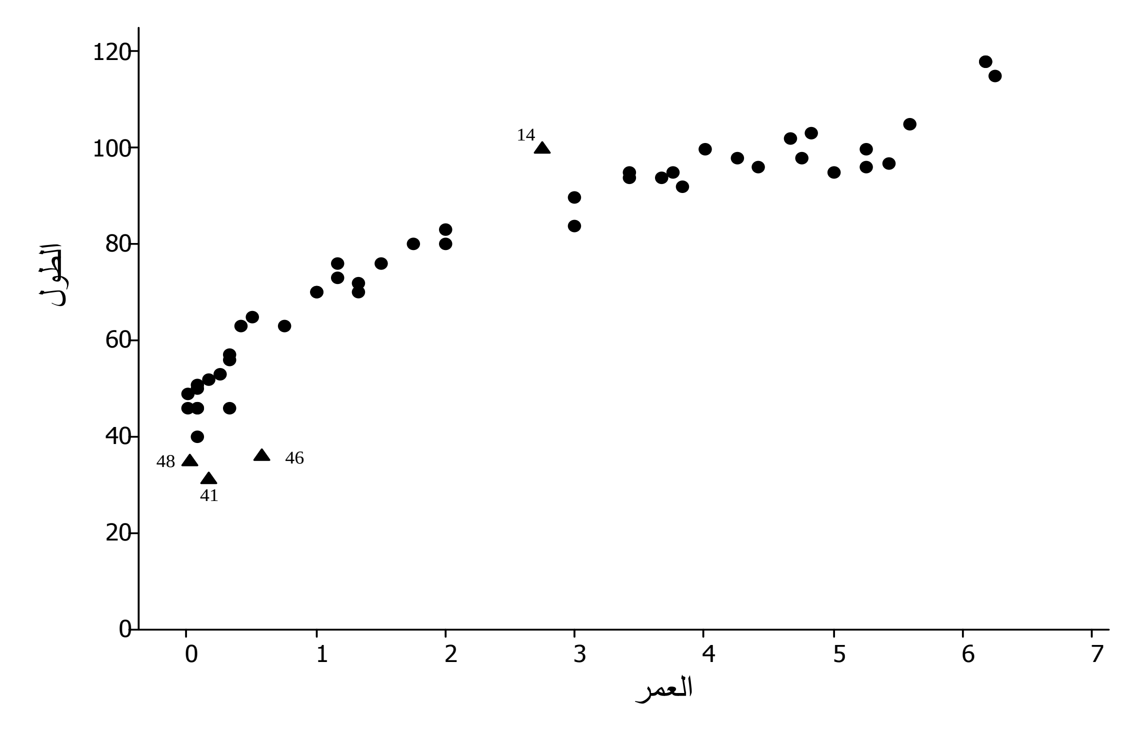
أولاً- تحديد المشاهدات الشاذة بحساب قيم الرافعة:

لقد تم توضيح الطريقة التي يتم بها حساب قيم الرافعات عند حساب البواقي المعيارية (انظر الفصل الثالث الجزء **٣-١٦-٢-٤**). ويعطي الجدول رقم (٤-١) بيانات نموذج الانحدار وقيم الرافعة h_{ii} لكل حالة. ويتضح من الجدول أن هناك أربع حالات تزيد قيم رافعاتها على ضعف متوسط قيم الرافعات $\left(\frac{2(p+1)}{n} = \frac{2(2+1)}{50} = 0.12\right)$ هي: (١٤) و(٤١) و(٤٦) و(٤٨). كما يلاحظ من الشكلين رقم (٤-١) و(٤-٢) أن نقاط هذه الحالات غير منسجمة مع بقية الحالات خاصة الحالتين (٤١) و(٤٦). وأما إذا تم الأخذ بالعتبة (Cut-off) التي حددها جون فوكس، فنجد أن الحالة رقم (٤١) هي الحالة الوحيدة التي تزيد قيمة رافعتها على ثلاثة أضعاف قيم الرافعات (٠,١٨). كما يلاحظ أنه لا توجد حالات شاذة حسب العتبة التي حددها نيتر وآخرون وهي (0.5).

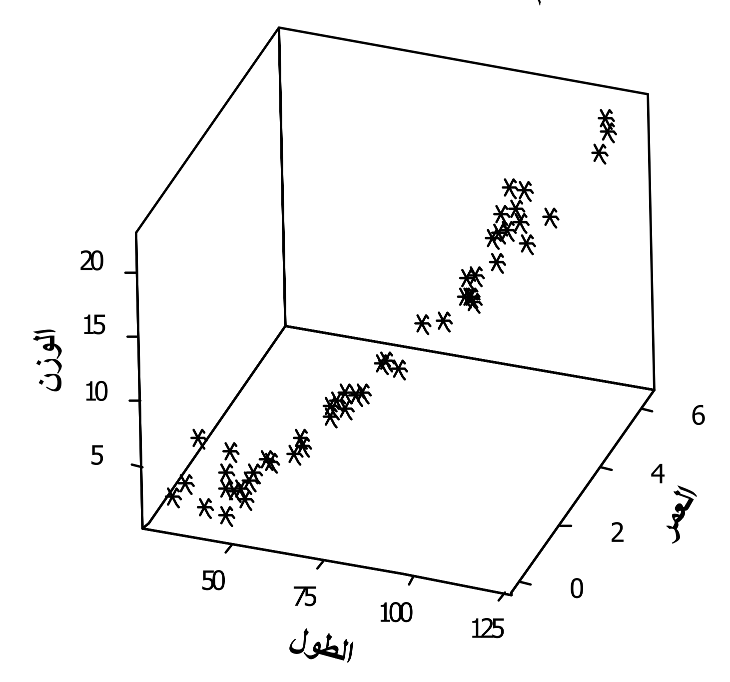
جدول رقم (٤-١): قيم الرافعة لنموذج انحدار وزن الطفل على متغيري العمر والطول

رقم المشاهدة	قيمة الرافعة h_{ii}	رقم المشاهدة	قيمة الرافعة h_{ii}
1	0.0221	26	0.0401
2	0.0795	27	0.0356
3	0.0555	28	0.0302
4	0.0426	29	0.0406
5	0.0304	30	0.0514
6	0.0465	31	0.0451
7	0.0877	32	0.0498
8	0.0397	33	0.0959
9	0.0315	34	0.0485
10	0.1061	35	0.0506
11	0.0684	36	0.0423
12	0.0417	37	0.0451
13	0.0503	38	0.0465
14	0.1260	39	0.0366
15	0.0932	40	0.0328
16	0.0445	41	0.1804
17	0.0372	42	0.0545
18	0.0357	43	0.0609
19	0.0518	44	0.0442
20	0.0784	45	0.0536
21	0.0360	46	0.1653
22	0.0877	47	0.0545
23	0.0344	48	0.1223
24	0.0746	49	0.0470
25	0.0392	50	0.0856

مصدر البيانات: مستشفى أبها للنساء والولادة والمراكز الصحية (١٩٩٩). بيانات عن أوزان ٥٠ طفلاً غير منشورة.



شكل رقم (٤-١): شكل انتشار طول وعمر الطفل



شكل رقم (٤-٢): شكل انتشار ثلاثي الأبعاد بين الوزن والطول وعمر الطفل

تحديد المشاهدات الشاذة بحساب مسافة مهالانوبس:

لحساب مسافة مهالانوبس حسب المعادلة (4.5) نحتاج حساب لمتجه $\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)$ و $\mathbf{S}^{-1}$ كما يلي:

متجه الوسط الحسابي:

$$\bar{\mathbf{x}} = \begin{pmatrix} 2.3472 \\ 76.400 \end{pmatrix}$$

ومصفوفة التباين والتغاير ومعكوسها:

$$\mathbf{S} = \begin{pmatrix} 4.4126 & 47.1681 \\ 47.1681 & 576.367 \end{pmatrix} ،\quad \mathbf{S}^{-1} = \begin{pmatrix} 1.80984 & -0.14811 \\ -0.14811 & 0.013856 \end{pmatrix}$$

فمسافة مهالانوبس للحالة الأولى $\mathrm{MD}_1 = \left(\mathbf{x}_1 - \bar{\mathbf{x}}\right)^{\mathrm{T}} \mathbf{S}^{-1} \left(\mathbf{x}_1 - \bar{\mathbf{x}}\right)$ مثلاً،

$$\mathrm{MD}_1 = \begin{pmatrix} 3.0 - 2.3472 \\ 84.0 - 76.40 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1.80984 & -0.14811 \\ -0.14811 & 0.013856 \end{pmatrix} \begin{pmatrix} 3.0 - 2.3472 \\ 84.0 - 76.40 \end{pmatrix} = 0.10194$$

كما يمكن الحصول على قيمة مسافة مهالانوبس للمشاهدة الأولى باستخدام المعادلة (٤,٧) كما يلي:

$$\mathrm{MD}_1 = (n-1)\left(h_{11} - \frac{1}{n}\right) = (50-1)\left(0.0220803 - \frac{1}{50}\right) = 0.10194$$

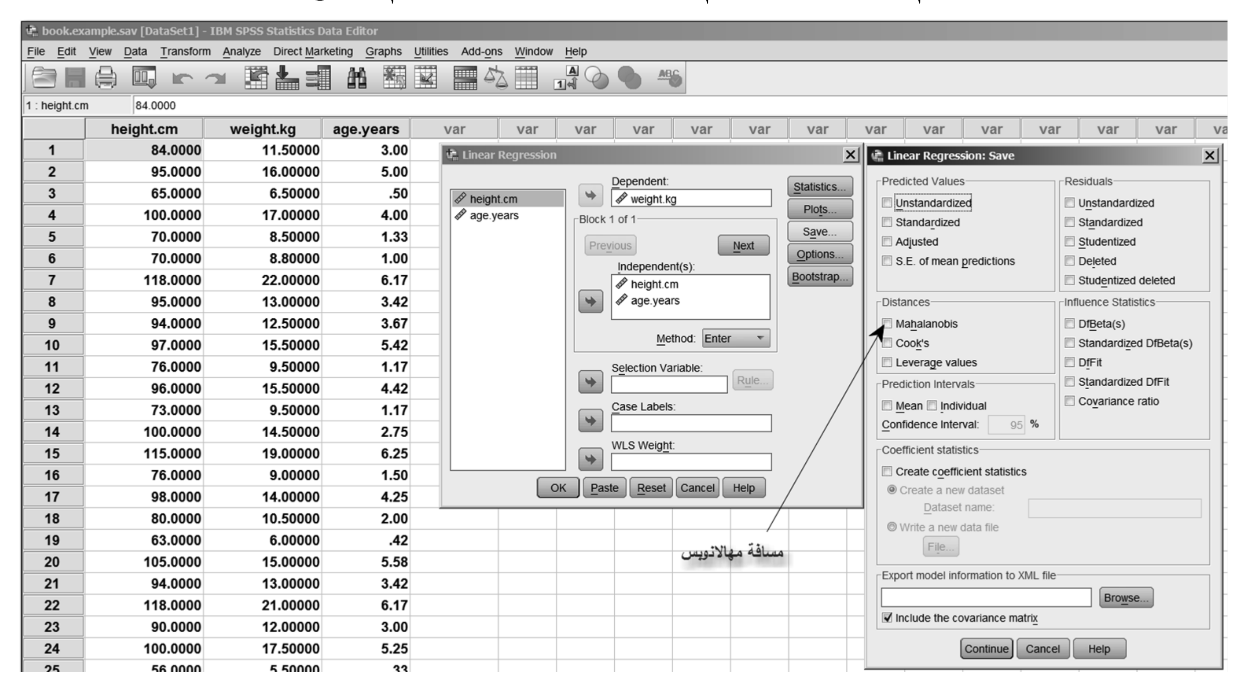
ويمكن الحصول على قيم مسافة مهالانوبس باستخدام برنامج **SPSS** باختيار **Analyze** و**Regression** و **Linear** ومن ثم تحديد المتغير التابع والمتغيرات المستقلة ومن خيار **Save** يتم اختيار **Mahalanobis** (الإطار رقم ٤-١).

ويوضح الجدول رقم (٤-٢): قيم مسافة مهالانوبس لنموذج انحدار وزن الطفل على متغيري العمر والطول. وبمقارنة قيم مسافة مهالانوبس بالقيمة الحرجة لتوزيع مربع كاي عند درجات حرية (٢) ومستوى دلالة (٠,٠٥) - ٥,٩٩-، فيتضح أن الحالتين (٤١) و(٤٦) والتي تزيد قيمة **مسافة مهالانوبس** لكل منهما شاذّتان.

جدول رقم (٤-٢): قيم مسافة مهالانوبس لنموذج انحدار وزن الطفل على متغيري العمر والطول

رقم الحالة	قيمة MD_i	رقم الحالة	قيمة MD_i
1	0.10194	26	0.98695
2	2.91383	27	0.76506
3	1.73828	28	0.49914
4	1.10675	29	1.01011
5	0.51174	30	1.53954
6	1.29825	31	1.23095
7	3.31934	32	1.46191
8	0.96570	33	3.71950
9	0.56243	34	1.39702
10	4.21779	35	1.50054
11	2.37081	36	1.09184
12	1.06427	37	1.22984
13	1.48263	38	1.29825
14	5.19496	39	0.81509
15	3.58663	40	0.62603
16	1.20084	41	7.85828
17	0.84253	42	1.69153
18	0.76800	43	2.00431
19	1.56010	44	1.18369
20	2.86005	45	1.64449
21	0.78190	46	7.11850
22	3.31934	47	1.69153
23	0.70417	48	5.01051
24	2.67431	49	1.32248
25	0.94088	50	3.21543

إطار رقم (٤-١): حساب قيم مسافة مهالانوبس باستخدام برنامج SPSS



٤-٢-٢ تحديد مشاهدات المتغير التابع الشاذة:

للكشف عن مشاهدات المتغير التابع الشاذة يستخدم عادة بواقي ستودنت المحذوفة (Studentized Deleted Residuals) والتي يتم الحصول عليها بإيجاد القيمة المعيارية للباقي المحذوف (Deleted Residual) والذي يسمى أيضاً بخطأ التنبؤ (Prediction Error). والباقي المحذوف للمشاهدة رقم (i) يساوي الفرق بين قيمة (y_i) الفعلية والقيمة المقدرة لها $\left(\hat{y}_{i(i)}\right)$باستخدام نموذج الانحدار الذي يتم تقديره بعد استبعاد المشاهدة رقم (i). وفيما يلي خطوات حساب الباقي المحذوف:

- يتم أولاً حذف الحالة رقم (i) ومن ثم يتم بناء نموذج انحدار باستخدام بقية الحالات (n-1).
- يتم تقدير القيمة المتوقعة للمتغير التابع رقم (i) - $\left(\hat{y}_{i(i)}\right)$ – وذلك بتعويض قيم المتغيرات المستقلة المناظرة للحالة رقم (i) ($x_{1i}, x_{2i},\ldots,x_{pi}$) في معادلة الانحدار التي تم تقديرها في الخطوة السابقة.
- يتم حساب الباقي المحذوف كما يلي:

$$d_i = y_i - \hat{y}_{i(i)} \qquad (4\text{-}8)$$

حيث إن d_i الباقي المحذوف رقم (i)، و y_i القيمة الفعلية للمتغير التابع رقم (i)، و $\hat{y}_{i(i)}$ القيمة المقدرة للمتغير التابع رقم (i) بعد حذف الحالة رقم (i).

ومن المعادلة (4.8) يتضح أننا نحتاج إلى بناء عدد (n) نموذج انحدار لحساب البواقي المحذوفة لجميع المشاهدات. ومن حسن الحظ توجد معادلة جبرية تمكننا من حساب البواقي المحذوفة دون الحاجة إلى بناء هذا العدد من النماذج (للبرهان انظر Montgomery, Peck and Vining, 2001, pp.598-600)، هي:

$$d_i = \frac{e_i}{1-h_{ii}}, \qquad i=1,2,\ldots,n \qquad (4\text{-}9)$$

ويلاحظ من المعادلة (4.9) أنه كلما كانت قيمة الرافعة (h_{ii}) كبيرة كانت قيمة الباقي المحذوف كبيرة. ويتم حساب باقي ستودنت المحذوف كما يلي:

$$d_i^* = \frac{d_i}{s.e(d_i)} \qquad (4\text{-}10)$$

حيث إن:

d_i^* = باقي ستودنت المحذوف رقم (i).

d_i = الباقي المحذوف رقم (i).

$s.e(d_i)$ = الانحراف المعياري للباقي المحذوف رقم (i).

ولباقي ستودنت المحذوف $\left(d_i^*\right)$ توزيع t بدرجات حرية (n-p-2) ولذلك جاءت التسمية. ويمكن حساب بواقي ستودنت دون الحاجة إلى بناء عدد (n) نموذج انحدار وذلك باستخدام المعادلة[1] التالية:

$$d_i^* = e_i\left[\frac{n-p-2}{RSS(1-h_{ii})-e_i^2}\right]^{\frac{1}{2}} \qquad (4\text{-}11)$$

حيث إن e_i الباقي العادي $\left(y_i-\hat{y}_i\right)$، وn عدد المشاهدات، وp عدد المتغيرات المستقلة، وRSS مجموع مربعات البواقي لنموذج الانحدار المقدر باستخدام جميع المشاهدات، وh_{ii} قيمة الرافعة i.

ولتحديد مشاهدات المتغير التابع (Y) الشاذة تتم مقارنة القيمة المطلقة لباقي ستودنت المحذوف بقيمة توزيع t عند درجات حرية (n-p-2) ومستوى معنوية (α)، فإذا كانت:

$$\left|d_i^*\right| > t_{\alpha,\, n-p-2}$$

تُعَدّ الحالة i (y_i) حالة شاذة تستدعي دراستها وتحديد مدى تأثيرها على مقدرات المربعات الصغرى.

---

[1] للبرهان يرجى الرجوع إلى **(Neter et al.1990 p. 400)**

مثال:

من المثال السابق (نموذج انحدار وزن الطفل) احسب البواقي المحذوفة وبواقي ستودنت المحذوفة ومن ثم حدد المشاهدات الشاذة في المتغير التابع؟

الحل:

باستخدام المعادلة (4.6) يتم حساب الباقي المحذوف للمشاهدتين الأوليين كما يلي:

$$d_1 = \frac{e_1}{1-h_{11}} = \frac{-0.38464}{(1-0.02208)} = -0.39332$$

$$d_2 = \frac{e_2}{1-h_{22}} = \frac{0.34346}{(1-0.079466)} = 0.37311$$

ويتم حساب بواقي ستودنت المحذوفة حسب المعادلة (4.8) على النحو التالي:

$$d_1^* = -0.38464\left(\frac{50\text{-}2\text{-}2}{56.6959(1-0.0220803)-(-0.38464)^2}\right)^{\frac{1}{2}} = -0.35082$$

$$d_2^* = 0.34346\left(\frac{50\text{-}2\text{-}2}{56.6959(1-0.0794658962)-(0.34346)^2}\right)^{\frac{1}{2}} = 0.32281$$

ويوضح الجدول رقم (٤-٣) قيم المتغير التابع، الباقي، الباقي المحذوف وباقي ستودنت المحذوف. ومن الجدول يتضح أن بواقي ستودنت المحذوفة التي تزيد قيمها على قيمة توزيع t ($t_{46;0.05}$ =2.013) هي للحالات: (٧)، (٢٠)، (٤٥) و(٤٦) والتي تُعدُّ حالات شاذة تستدعي دراستها وقياس أثرها على مقدرات المربعات الصغرى. وكما يجب ملاحظة أن لهذه الحالات أكبر قيم مطلقة للبواقي، بلغت على التوالي: ٢,٠٧٣٣٥، -٢,٥٩٨٧٤، -٢,٢١٠٤٧، و٢,٥٠٠٧٩. وكذلك يلاحظ من الجدول نفسه أن الحالة رقم (٤٦) حالة شاذة في المتغير التابع وفي المتغيرين المستقلين.

جدول رقم (٤-٣): بواقي ستودنت المحذوف لنموذج انحدار الطفل على متغيري العمر والطول

رقم المشاهدة	وزن الطفل (كجم)	الباقي (e_i)	الباقي المحذوف (d_i)	بواقي ستودنت (d'_i)
1	11.5	-0.38464	-0.39332	-0.35082
2	16	0.34346	0.37311	0.32281
3	6.5	-0.01575	-0.01667	-0.01459
4	17	1.92140	2.00687	1.83217
5	8.5	0.36472	0.37617	0.33405
6	8.8	1.06099	1.11272	0.98906
7	22	2.07335	2.27276	2.04201
8	13	-0.75927	-0.79067	-0.70163
9	12.5	-1.43490	-1.48154	-1.33866
10	15.5	-0.91003	-1.01802	-0.87415
11	9.5	0.80942	0.86883	0.76009
12	15.5	0.41535	0.43343	0.38279
13	9.5	1.18313	1.24574	1.10804
14	14.5	0.92240	1.05541	0.89647
15	19	-0.64900	-0.71570	-0.61642
16	9	-0.08685	-0.09089	-0.08004
17	14	-1.12966	-1.17330	-1.04934
18	10.5	0.31446	0.32609	0.28870
19	6	-0.17054	-0.17986	-0.15780
20	15	-2.59874	-2.81971	-2.61299
21	13	-0.63470	-0.65837	-0.58443
22	21	1.07335	1.17658	1.02371
23	12	-0.63207	-0.65457	-0.58151
24	17.5	0.92039	0.99457	0.86884
25	5.5	0.30954	0.32217	0.28470
26	5.3	-0.01503	-0.01566	-0.01382
27	6.5	-0.06680	-0.06927	-0.06128
28	13.5	-0.37788	-0.38965	-0.34608
29	4.5	-0.22068	-0.23002	-0.20303
30	15.5	-0.23006	-0.24253	-0.21287
31	16.5	0.36771	0.38509	0.33938
32	11	1.11466	1.17312	1.04211
33	17.5	1.41868	1.56918	1.37114
34	14.55	-1.89899	-1.99581	-1.81529
35	10	-0.55926	-0.58908	-0.51852
36	4	-0.50004	-0.52211	-0.46131
37	3.5	-0.64282	-0.67318	-0.59481
38	8	0.26099	0.27371	0.24090
39	8	-0.38442	-0.39904	-0.35327
40	14	-0.15554	-0.16081	-0.14249
41	1.75	-0.13402	-0.16351	-0.13336
42	3.2	-0.44453	-0.47017	-0.41255
43	5.55	1.60527	1.70938	1.52957
44	2.75	-1.51739	-1.58749	-1.42868
45	1.35	-2.21047	-2.33557	-2.14672
46	5.5	2.50079	2.99595	2.64659
47	4.5	0.85547	0.90480	0.79793
48	3.25	1.04781	1.19376	1.01870
49	3.3	-0.62218	-0.65286	-0.57615
50	1.4	-1.49710	-1.63728	-1.44174

٤-٣ تحديد الحالات المؤثرة (Identifying Influential Cases):

إن الخطوة التالية للكشف عن الحالات الشاذة في المتغيرات المستقلة والمتغير التابع هي تحديد ما إذا كانت هذه الحالات مؤثرة (Influential) أم لا؟. وتعتبر الحالة مؤثرة إذا كان استبعادها يحدث تغيراً ملحوظاً في قيم معاملات نموذج الانحدار والإحصاءات المرتبطة بها. وتوجد أربعة مقاييس تستخدم لتحديد الحالات المؤثرة هي: ,DFITS , DFBETAS، Cook's Distance وCOVRATIO. وتعتمد هذه المقاييس في تحديد أثر الحالة الشاذة على قياس الفرق بين قيم مقدرات المربعات الصغرى باستخدام كل الحالات (n) وبإسقاط حالة واحدة (n-1).

٤-٣-١ مقياس DFITS:

يستخدم مقياس DFFITS لقياس أثر الحالة رقم (i) على القيمة الموفقة أو المقدرة $(\hat{y}_i)$. ولقد طور بيلسلي وآخرون (Belsley et al (1980) pp14-15) المعادلة التالية لقياس أثر الحالة (i)على القيمة الموفقة$(\hat{y}_i)$:

$$DFITS_i = d_i^* \left(\frac{h_{ii}}{1-h_{ii}} \right)^{\frac{1}{2}} \qquad (4\text{-}12)$$

حيث إن:

(d_i^*) = باقي ستودنت المحذوف رقم (i).

h_{ii} = قيمة الرافعة وهي العنصر القطري رقم i من مصفوفة القبعة $\mathbf{x}^T$ $\mathbf{H} = \mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}$.

وحسب بيلسلي وآخرين (Belsley et al (1980) p28) تعتبر الحالة مؤثرة إذا كانت قيمة DFITS المطلقة أكبر من $2\sqrt{\frac{(p+1)}{n}}$، أي:

$$\left|DFITS_i\right| \; > \; 2\sqrt{\frac{(p+1)}{n}} \qquad (4\text{-}13)$$

ويقترح نيتر وآخرون (Neter et al (1990) p401) أنه تعتبر الحالة مؤثرة إذا كانت القيمة المطلقة لـ DFITS أكبر من واحد صحيح في حالة العينات الصغيرة والمتوسطة وإذا كانت أكبر من $2\sqrt{\frac{(p+1)}{n}}$ في حالة العينات الكبيرة. وأما شاترجي وهادي ((Chatterjee & Hadi (1988) فيقترحان مقارنة القيمة المطلقة لـ DFITS بقيمة أكبر قليلاً من تلك التي اقترحها بيلسلي وزملاؤه، هي:

$$\left|DFITS_i\right| \; > \; 2\sqrt{\frac{(p+1)}{n-p-1}} \qquad (4\text{-}14)$$

٤-٣-٢ قياس الأثر على معاملات الانحدار (مقياس (DFBETAS)):

يستخدم مقياس DFBETAS الذي طوره بيلسلي وآخرون (Belsley et al (1980) p13) لقياس الفرق بين قيم معاملات الانحدار المقدرة باستخدام جميع المشاهدات (n) وقيم معاملات الانحدار المقدرة بعد إسقاط الحالة رقم (i)، أي باستخدام (n-1) مشاهدة. ويتم الحصول على الفرق بين قيمتي معامل الانحدار باستخدام كل الحالات (n) وباستخدام (n-1) حالة بعد اسقاط الحالة رقم (i) باستعمال المعادلة التالية:

$$\text{DFBETAS}_{k(i)} = \frac{\hat{\beta}_k - \hat{\beta}_{k(i)}}{S_{(i)}\sqrt{\left(\mathbf{X}^T\mathbf{X}\right)^{-1}_{kk}}}, \quad \text{for} \quad k = 0,1,\ldots,p \qquad (4\text{-}15)$$

حيث إن:

$\hat{\beta}_k$ = معامل الانحدار رقم k المقدر باستخدام كل الحالات (n).

$\hat{\beta}_{k(i)}$ = معامل الانحدار رقم k المقدر باستخدام (n-1) حالة بعد إسقاط الحالة رقم (i).

$S_{(i)}$ = الخطأ المعياري للتقدير لنموذج الانحدار الموفق باستخدام (n-1) حالة، أي بعد إسقاط الحالة رقم (i).

$\left(\mathbf{X}^T\mathbf{X}\right)^{-1}_{kk}$ العنصر القطري رقم (k) من المصفوفة $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$.

كما يمكن الحصول على DFBETAS دون تكرار بناء (n) نموذج انحدار خطي، وذلك باستخدام المعادلة التالية (للبرهان انظر Montgomery, Peck and Vining, 2001, pp. 609):

$$\text{DFBETAS}_{k(i)} = \frac{r_{k,i}}{\sqrt{\mathbf{r}_k^T\mathbf{r}_k}}\frac{d_i^*}{\sqrt{1-h_{ii}}} \quad \text{for} \quad k = 0,1,\ldots,p \qquad (4\text{-}16)$$

حيث إن:

$r_{k,i}$ = هي القيمة رقم (k,i) من المصفوفة $\mathbf{R} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$

$\mathbf{r}_k^T\mathbf{r}_k$ = هي القيمة رقم (kk) من المصفوفة $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$.

d_i^* = باقي ستودنت للمشاهدة رقم (i)،

قيم وتشير القيمة المطلقة الكبيرة لـ DFBETAS إلى أن الحالة رقم (i) حالة مؤثرة على قيمة معامل الانحدار رقم (k). وكمعيار عام لتحديد الحالات المؤثرة، يقترح نيتر وآخرون (Neter et al (1990) p403) إذا كانت قيمة DFBETAS أكبر من واحد صحيح في حالات العينات الصغيرة أو أكبر من $2/\sqrt{n}$ في حالة العينات الكبيرة تعتبر الحالة مؤثرة، أي تعتبر الحالة رقم (i) مؤثرة إذا تحقق الشرط التالي:

في حالة العينات الصغيرة: $\left|DFBETAS_{K(i)}\right| > 1$

أو في حالة العينات الكبيرة: $\left|DFBETAS_{K(i)}\right| > 2/\sqrt{n}$

٤-٣-٣ قياس الأثر على كل معاملات الانحدار (مقياس كوك[٢] (Cook's Distance Measure)):

يستخدم مقياس كوك لقياس أثر الحالة رقم (i) على كل معاملات نموذج الانحدار. ويختلف مقياس كوك عن مقياس DFBETAS في أن الأول يقيس أثر إسقاط الحالة رقم (i) على كل قيم معاملات نموذج الانحدار، في حين يقيس الثاني أثر إسقاط الحالة رقم (i) على كل معامل من معاملات النموذج على حده. ويأخذ مقياس كوك الصيغة التالية:

$$D_i = \frac{\left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}\right)^T \mathbf{x}^T\mathbf{x}\left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}\right)}{(p+1)\times S^2} \qquad (4\text{-}17)$$

حيث إن:

$\hat{\boldsymbol{\beta}}$ = متجه معاملات نموذج الانحدار باستخدام كل الحالات (n).

$\hat{\boldsymbol{\beta}}_{(i)}$= متجه معاملات نموذج الانحدار بعد إسقاط الحالة رقم (i).

$\mathbf{x}$= مصفوفة البيانات من الدرجة ($n \times (p+1)$).

p = عدد المتغيرات المستقلة.

S^2= تباين نموذج الانحدار باستخدام كل الحالات.

ويمكن حساب D_i حسب المعادلة التالية دون الحاجة إلى بناء عدد (n) نموذج انحدار خطي كما يلي:

$$D_i = \frac{e_i^2}{(p+1)\times S^2}\left(\frac{h_{ii}}{(1-h_{ii})^2}\right) \qquad (4\text{-}18)$$

ويلاحظ من المعادلة (4.18) أن قيمة D_i تعتمد على قيمتي الباقي (e_i) وقيمة الرافعة (h_{ii}). فإذا كانت قيمة أي من (e_i) و(h_{ii}) كبيرة تكون قيمة D_i أيضا كبيرة. ولتحديد أثر الحالة (i) على معاملات نموذج الانحدار يقترح فوكس (Fox (1993) pp276) أن تتم مقارنة قيمة D_i بالقيمة $\left(\frac{4}{(n-p-1)}\right)$، فإذا كانت قيمة D_i أكبر من هذه القيمة فتُعَدُّ الحالة (i) حالة

[٢] Cook, R. D. (1977), Detection of Influential Observations in Linear Regression. *Technometrics 19: pp15-18*

مؤثرة على قيم معاملات نموذج الانحدار وإلا فإنها تُعَدَّ غير مؤثرة. كما يمكن استخدام وسيط توزيع F لاختبار فرض أن قيمة D_i أكبر من الصفر، أي اختبار الفرض التالي:

$$H_0: D_i = 0 \text{ مقابل } H_0: D_i \neq 0$$

فإذا كانت قيمة D_i أكبر من القيمة الحرجة لوسيط توزيع F $F_{0.5,(p+1),(n-p-1)}$ أو أكبر من الواحد الصحيح، فتعد الحالة مؤثرة (Weisberg, 2005, pp.199-200) .

٤-٣-٤ الأثر على الأخطاء المعيارية (Influence on Standard Errors):

يُستخدم مقياس **COVRATIO** الذي طوّره بيلسلي وآخرون (Belesley et al. (1980) (pp.22-24) لقياس أثر أي حالة على مصفوفة تباين وتغاير معاملات الانحدار المقدرة. وهذا المقياس عبارة عن نسبة محدّدة مصفوفة تباين-تغاير معاملات الانحدار المقدرة بعد حذف الحالة رقم i لمحدّدة مصفوفة تباين-تغاير معاملات الانحدار المقدر باستخدام جميع الحالات (**n**)؛ أي أن:

$$\text{COVRATIO}_i = \frac{\det\left(S_{(i)}^2(\mathbf{x}_{(i)}^T\mathbf{x}_{(i)})^{-1}\right)}{\det\left(S^2(\mathbf{x}^T\mathbf{x})^{-1}\right)} \qquad (4\text{-}19)$$

حيث إن:

$S_{(i)}^2$**:** التباين المقدر بعد حذف الحالة رقم i

$\mathbf{x}_{(i)}$**:** مصفوفة البيانات بعد حذف المشاهدة رقم i من الدرجة $(n-1)\times(p+1)$.

وبإعادة تنظيم المعادلة[٣] (4.19) يمكن الحصول على:

$$\text{COVRATIO}_i = \frac{1}{(1-h_{ii})\left(\dfrac{n-p-2+d_i^{*2}}{n-p-1}\right)^{p+1}} \qquad (4\text{-}20)$$

ويلاحظ من الصيغة (4.20) أن قيمة COVRATIO تعتمد على قيمتي القبعة/الرافعة (h_{ii}) وباقي ستودنت $\left(d_i^*\right)$. حيث تزيد قيمة COVRATIO بزيادة قيمة h_{ii} وبانخفاض قيمة بواقي ستودنت المحذوفة $\left(d_i^*\right)$. ولتحديد أثر الحالة رقم i على الأخطاء المعيارية لمعاملات الانحدار المقدرة، يقترح بيلسلي وآخرون أن تتم مقارنة قيمة COVRATIO المطلقة بالقيمة (1+3(p+1)/n)؛ فإذا كانت قيمة COVRATIO المطلقة أكبر من هذه العتبة فإن الحالة i تُعدَّ مؤثرة على قيم الأخطاء المعيارية لمعاملات نموذج الانحدار.

---

[٣] للبرهان انظر بيلسلي وآخرين (Belsley et al. (1980) pp. 22-24)

مثال:

من المثال السابق (نموذج انحدار وزن الطفل على متغيري العمر والطول)، احسب قيم DFITS وDFBETAS و Cook's distance وCOVRATIO، ثم حدد الحالات المؤثرة على القيم الموفقة ومعاملات نموذج الانحدار كل على حدة وعلى كل معاملات الانحدار وعلى الأخطاء المعيارية لمعاملات الانحدار المقدرة؟.

الحل:

- **حساب قيم DFITS وتحديد أثر الحالات المؤثرة على القيم الموفقة:**

باستخدام قيم بواقي ستودنت المحذوفة والرافعات ، يتم حساب قيم DFITS حسب المعادلة (4.12) للحالتين الأوليين كما يلي:

$$DFITS_1 = d_1^*\left(\frac{h_{11}}{1-h_{11}}\right)^{\frac{1}{2}} = -0.3508182\times\sqrt{\left(\frac{0.0220803}{1-0.0220803}\right)} = 0.05271486$$

$$DFITS_2 = d_2^*\left(\frac{h_{22}}{1-h_{22}}\right)^{\frac{1}{2}} = 0.3228095\times\sqrt{\left(\frac{0.0794659}{1-0.0794659}\right)} = 0.0948455$$

ويستعرض الجدول رقم (٤-٤) قيم DFITS لجميع الحالات. وحسب معيار بيلسلي تُعَدًّ الحالة مؤثرة إذا كانت القيمة المطلقة لـ DFITS أكبر من:

$$2\sqrt{\frac{p+1}{n}} = 2\times\sqrt{\frac{2+1}{50}} = 0.4899$$

ووفقاً لهذا الحد نجد أن الحالات المؤثرة على قيم المتغير التابع الموفقة ($\hat{y}_i$) هي: (٧)، (٢٠)، (٤٥) و(٤٦). ولتحديد الحالات المؤثرة وفقاً لمعيار شاترجي وهادي يتم حساب القيمة التالية:

$$2\sqrt{\frac{p+1}{n-p-1}} = \sqrt{\frac{2+1}{50-2-1}} = 0.50529$$

وعلى الرغم من الزيادة في قيمة العتبة التي تحدد الحالات المؤثرة حسب معيار شاترجي وهادي لم تتغير الحالات المؤثرة.

- **حساب قيم DFBETAS وتحديد أثر الحالات المؤثرة على قيم معاملات النموذج:**

لحساب DFBETAS يتخذ الحل الخطوات التالية:

١) بناء عدد (٥٠) نموذج انحدار من (٤٩) حالة، بحذف حالة واحدة في كل مرة. فمثلاً نبدأ بإسقاط الحالة الأولى ويتم تقدير معالم النموذج والخطأ المعياري من بقية الحالات البالغة (٤٩) حالة. فمثلاً لحساب قيم DFBETAS للمشاهدة الأولى، يتم إسقاط المشاهدة الأولى ومن ثم بناء النموذج للحصول على القيم التالية:

جدول (٤-٤) نتائج نموذج انحدار وزان الطفل على عمره وطوله بعد إسقاط المشاهدة الأولى

ANOVA:

	df	SS	MS	F	Significance F
Regression	2	1441.503	720.7516	597.492	0.000
Residual	47	56.696	1.206296		
Total	49	1498.199			
Predictor	Coefficients	Standard Error	t Stat	P-value	
Intercept	-2.18187	0.976	-2.23513	0.03020	
Age	1.20080	0.211	5.68883	0.00000	
Height	0.12457	0.018	6.74488	0.00000	

٢) حساب معكوس المصفوفة $\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)$:

$$\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}=\begin{pmatrix} 0.789947 & 0.144238 & -0.0145092 \\ 0.144238 & 0.036936 & -0.0030227 \\ -0.014509 & -0.003023 & 0.0002828 \end{pmatrix}$$

٣) باستخدام المعادلة رقم (4.15) يتم حساب قيم DFBETAS. فمثلاً للحالة الأولى يتم حساب المصفوفة $\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}$ وبناء نموذج انحدار جميع المشاهدات عدا المشاهدة الأولى للحصول على مقدرات معاملات الانحدار والتباين ومن ثم نجد أن:

$$\text{DFBETAS }\widehat{\beta}_{0(1)}=\frac{\widehat{\beta}_0-\widehat{\beta}_{0(1)}}{S_{(1)}\sqrt{\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}_{00}}}=\frac{(-2.18186891)-(-2.180339335)}{1.10870682\sqrt{0.789947}}=-0.00155$$

$$\text{DFBETAS }\widehat{\beta}_{1(1)}=\frac{\widehat{\beta}_1-\widehat{\beta}_{1(1)}}{S_{(1)}\sqrt{\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}_{11}}}=\frac{1.20080469-1.201252723}{1.10870682\sqrt{0.036936}}=-0.00210$$

$$\text{DFBETAS}\ \hat{\beta}_{2(1)} = \frac{\hat{\beta}_2 - \hat{\beta}_{2(1)}}{S_{(1)}\sqrt{\left(\mathbf{x}^T\mathbf{x}\right)^{-1}_{22}}} = \frac{0.124572515 - 0.124641693}{1.10870682\sqrt{0.0002828}} = -0.00371$$

وباستخدام المعادلة (4.16) يتم الحصول على نفس قيم DFBETAS. ولحساب $r_{k,i}$ نحتاج حساب المصفوفة $\mathbf{R} = \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T$، كما يلي:

$$\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T = \begin{pmatrix} 0.789947 & 0.144238 & -0.0145092 \\ 0.144238 & 0.036936 & -0.0030227 \\ -0.014509 & -0.003023 & 0.0002828 \end{pmatrix}\begin{pmatrix} 11.5 & 16 & ... & 1.4 \\ 3 & 5 & ... & 0.08 \\ 84 & 95 & ... & 40 \end{pmatrix} = \begin{pmatrix} 0.003889 & 0.132764 & ... & 0.221118 \\ 0.0011391 & 0.041761 & ... & 0.026286 \\ 0.0001759 & -0.00276 & ... & -0.00344 \end{pmatrix}$$

ومن ثم يتم حساب قيم DFBETAS للحالة أو المشاهدة الأولى كما يلي:

$$\text{DFBETAS}\ \hat{\beta}_{0(1)} = \frac{r_{0,1}}{\sqrt{\mathbf{r}_0^T\mathbf{r}_0}}\frac{d_1^*}{\sqrt{1-h_{11}}} = \frac{0.003889}{\sqrt{0.789947}} \times \frac{-0.35082}{\sqrt{1-0.0220803}} = -0.00155$$

$$\text{DFBETAS}\ \hat{\beta}_{1(1)} = \frac{r_{1,1}}{\sqrt{\mathbf{r}_1^T\mathbf{r}_1}}\frac{d_1^*}{\sqrt{1-h_{11}}} = \frac{0.0011391}{\sqrt{0.036936}} \times \frac{-0.35082}{\sqrt{1-0.0220803}} = -0.00210$$

$$\text{DFBETAS}\ \hat{\beta}_{2(1)} = \frac{r_{2,1}}{\sqrt{\mathbf{r}_2^T\mathbf{r}_2}}\frac{d_1^*}{\sqrt{1-h_{11}}} = \frac{0.0001759}{\sqrt{0.000283}} \times \frac{-0.35082}{\sqrt{1-0.0220803}} = -0.00371$$

والجدول رقم (٤-٥) يعطي قيم DFBETAS لكل الحالات وباستخدام معيار $\left(2/\sqrt{n}\right)$ نجد أن هناك (٨) حالات مؤثرة على قيم معاملات نموذج انحدار الطفل على متغيري العمر والطول. ويلاحظ أن بعض هذه الحالات تؤثر على معامل واحد وبعضها يؤثر على معاملين، والبعض الآخر يؤثر على المعاملات الثلاث. وكما يلاحظ من الجدول رقم (٤-٥) أن هناك ثلاث حالات – ٣٣، ٤٣، ٤٥ و٥٠- ظهرت كحالات مؤثرة على قيم معاملات نموذج الانحدار ولم تظهر كحالات شاذة سواء في المتغير التابع أوفي المتغيرين المستقلين.

جدول رقم (٤-٥): الحالات المؤثرة على قيم معاملات نموذج الانحدار:

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
14		14
	20	
	33	
43		
45		
46	46	46
48		48
50		50

وأخذاً بمعيار نيتر وآخرين الذي يحدد الحالات المؤثرة إذا كانت قيم DFBETAS أكبر من واحد صحيح في حالة العينات الصغيرة والمتوسطة، نجد أن الحالة (٤٦) هي الحالة الوحيدة المؤثرة على معاملي النموذج $\hat{\beta}_0$ و $\hat{\beta}_2$ أي المعامل الثابت ومعامل الطول على التوالي.

- **حساب قيم Cook's Distance وتحديد أثر الحالات المؤثرة على قيم معاملات نموذج الانحدار مجتمعة:**

باستخدام قيم البواقي العادية والرافعات والتباين يتم حساب قيم D_i حسب المعادلة (٤-١٨) للحالتين الأوليين كما يلي:

$$D_1 = \frac{e_1^2}{(p+1)\times S^2}\times\frac{h_{11}}{(1-h_{11})^2}=\frac{(-0.38464)^2}{1.2063\times 3}\times\frac{0.0220803}{(1-0.0220803)^2} = 0.000944$$

$$D_2 = \frac{e_2^2}{(p+1)\times S^2}\times\frac{h_{22}}{(1-h_{22})^2}=\frac{(0.34346)^2}{1.2063\times 3}\times\frac{0.079466}{(1-0.079466)^2} = 0.003057$$

ويعطي الجدول رقم (٤-٦) قيم D_i لكل الحالات. ووفقاً للحد الذي طوره فوكس (Fox (1993) p.276) أنه تتم مقارنة قيمة D_i بالقيمة ((4/(n-p-1))، أي:

$$4/(n-p-1) = 4/(50-2-1) = 0.085106$$

ويتضح من الجدول أن الحالات المؤثرة على كل معاملات الانحدار هي: الحالة (٧)، (٢٠)، و(٤٦).

وباستخدام وسيط توزيع F، فنجد أن القيمة الحرجة $\left(F_{0.5,3,47} = 0.800229\right)$ أكبر من جميع قيم D_i، مما يشير إلى عدم وجود حالات مؤثرة.

جدول رقم (٤-٦): قيم Di، DFBETAS، وDFFITS

D_i	DFBETAS $\hat{\beta}_2$	DFBETAS $\hat{\beta}_1$	DFBETAS $\hat{\beta}_0$	$DFFITS_i$	م
0.000944	-0.003710	-0.002100	-0.001550	-0.052715	1
0.003057	-0.055200	0.073110	0.050260	0.094845	2
0.000004	-0.002110	0.002640	0.001370	-0.003537	3
0.047395	0.186800	-0.100240	-0.177010	0.386415	4
0.001190	0.025520	-0.032170	-0.012920	0.059193	5
0.015908	0.136270	-0.160290	-0.092830	0.218405	6
0.125239	0.026490	0.171910	-0.077430	0.633290	7
0.006860	-0.085880	0.061830	0.076640	-0.142675	8
0.019092	-0.079150	0.030720	0.068200	-0.241335	9
0.030378	0.190400	-0.246450	-0.170940	-0.301125	10
0.014264	0.161340	-0.173210	-0.127580	0.205933	11
0.002166	-0.016810	0.035230	0.015220	0.079871	12
0.021552	0.175580	-0.196430	-0.128520	0.254890	13
0.038788	0.311120	-0.281700	-0.285180	0.340410	14
0.013191	0.033950	-0.092550	-0.016660	-0.197615	15
0.000102	-0.011920	0.012820	0.008880	-0.017274	16
0.014149	-0.022660	-0.027770	0.022790	-0.206247	17
0.001048	0.036150	-0.036260	-0.027230	0.055528	18
0.000463	-0.019620	0.025870	0.011590	-0.036897	19
0.172177	0.272630	-0.466740	-0.218440	-0.761957	20
0.004307	-0.061380	0.042040	0.053990	-0.112869	21
0.033564	0.013280	0.086180	-0.038820	0.317483	22
0.004069	-0.065900	0.052340	0.055370	-0.109710	23
0.020385	-0.112830	0.168620	0.097840	0.246647	24
0.001124	0.005680	-0.019410	0.008180	0.057508	25
0.000003	-0.000510	0.001160	-0.000170	-0.002826	26
0.000047	-0.003850	0.006000	0.001120	-0.011775	27
0.001266	0.001480	-0.013920	-0.002980	-0.061058	28
0.000594	0.003420	0.007260	-0.013300	-0.041773	29
0.000836	0.015010	-0.026680	-0.013080	-0.049562	30
0.001849	0.004500	0.015200	-0.006410	0.073773	31
0.018952	0.179480	-0.183240	-0.142390	0.238660	32
0.065257	-0.277140	0.359950	0.250370	0.446583	33
0.053395	-0.001890	-0.109420	0.016390	-0.409886	34
0.004854	-0.092270	0.090750	0.075350	-0.119734	35
0.003185	0.008930	0.016340	-0.031820	-0.096929	36
0.005647	0.022160	0.012480	-0.052070	-0.129265	37
0.000963	0.033190	-0.039040	-0.022610	0.053196	38
0.001612	-0.039180	0.045450	0.025460	-0.068890	39
0.000234	-0.008780	0.003320	0.007750	-0.026229	40
0.001333	0.054810	-0.043550	-0.060440	-0.062562	41
0.003330	0.043990	-0.017990	-0.064000	-0.099069	42
0.049175	-0.234560	0.142770	0.302120	0.389526	43
0.030750	0.028630	0.052950	-0.101140	-0.307074	44
0.080735	0.201000	-0.063880	-0.307770	-0.510685	45
0.409923	-1.047780	0.856800	1.144880	1.177657	46
0.012334	-0.085070	0.034800	0.123780	0.191612	47
0.048142	-0.302130	0.221690	0.348680	0.380187	48
0.005534	0.022920	0.011890	-0.052460	-0.127934	49
0.063424	0.308430	-0.206210	-0.375100	-0.441178	50

- حساب قيم COVRATIO وتحديد أثر الحالات المؤثرة على القيم المعيارية:

باستخدام الصيغة (4.20)، يمكن حساب قيمتي COVRATIO للحالة الأولى كما يلي:

$$\text{COVRATIO}_1 = \frac{1}{(1-h_{11})\left(\frac{n-p-2+d_1^{*2}}{n-p-1}\right)^{p+1}} = \frac{1}{(1-0.0220803)\left(\frac{50-2-2+-0.3508182^2}{50-2-1}\right)^{2+1}}$$
$$=1.0820$$

ويعطي الجدول رقم (٤-٧) قيم COVRATIO لكل الحالات. وحسب معيار بيلسلي تعدًّ الحالة مؤثرة إذا كانت قيمة COVRATIO واقعة خارج المدى:

$$1-3(p+1)/n < \text{COVRATIO} < 1+3(p+1)/n$$

$$0.82 < \text{COVARIO} < 1.18$$

وطبقاً لهذا الحد نجد أن الحالات المؤثرة على الأخطاء المعيارية هي الحالتان: (٢٠) و (٤١). ويلاحظ أن هاتين الحالتين قد تم تحديدهما كمشاهدات شاذة؛ فالحالة (٤١) تم تحديدها حالة شاذة في المتغيرات المستقلة لكبر قيمة رافعاتها البالغة (٠,١٨٠٤) أما الحالة (٢٠) فقد تم تحديدها كحالة شاذة في المتغير التابع لصغر باقيها المحذوف البالغ (-٢,٦١٣٠).

جدول رقم (٤-٧): قيم بواقي ستودنت d_i^*، وقيم الرافعة (h_{ii})، وقيم COVRATIO

المشاهد	باقي ستودنت d_i^*	قيمة الرافعة h_{ii}	قيمة $COVRATIO_i$
1	-0.3508	0.0221	1.0820
2	0.3228	0.0795	1.1509
3	-0.0146	0.0555	1.1293
4	1.8322	0.0426	0.9019
5	0.3340	0.0304	1.0922
6	0.9891	0.0465	1.0502
7	2.0420	0.0877	0.9013
8	-0.7016	0.0397	1.0758
9	-1.3387	0.0315	0.9820
10	-0.8741	0.1061	1.1357
11	0.7601	0.0684	1.1029
12	0.3828	0.0417	1.1025
13	1.1080	0.0503	1.0378
14	0.8965	0.1260	1.1587
15	-0.6164	0.0932	1.1476
16	-0.0800	0.0445	1.1159
17	-1.0493	0.0372	1.0320
18	0.2887	0.0357	1.1001
19	-0.1578	0.0518	1.1231
20	-2.6130	0.0784	0.7641
21	-0.5844	0.0360	1.0822
22	1.0237	0.0877	1.0928
23	-0.5815	0.0344	1.0806
24	0.8688	0.0746	1.0977
25	0.2847	0.0392	1.1043
26	-0.0138	0.0401	1.1112
27	-0.0613	0.0356	1.1058
28	-0.3461	0.0302	1.0913
29	-0.2030	0.0406	1.1088
30	-0.2129	0.0514	1.1212
31	0.3394	0.0451	1.1087
32	1.0421	0.0498	1.0467
33	1.3711	0.0959	1.0462
34	-1.8153	0.0485	0.9109
35	-0.5185	0.0506	1.1041
36	-0.4613	0.0423	1.0984
37	-0.5948	0.0451	1.0916
38	0.2409	0.0465	1.1144
39	-0.3533	0.0366	1.0982
40	-0.1425	0.0328	1.1013
41	-0.1334	0.1804	1.2999
42	-0.4126	0.0545	1.1157
43	1.5296	0.0609	0.9788
44	-1.4287	0.0442	0.9796
45	-2.1467	0.0536	0.8463
46	2.6466	0.1653	0.8353
47	0.7979	0.0545	1.0826
48	1.0187	0.1223	1.1365
49	-0.5761	0.0470	1.0954
50	-1.4417	0.0856	1.0217

٤-٤ بعض الحلول المقترحة لمعالجة مشكلة وجود البيانات الشاذة:

إن وجود بيانات شاذة في مشاهدات المتغير التابع و/أو في مشاهدات المتغيرات المستقلة يستلزم دراستها لتحديد مصادرها وأسباب وجودها. فإذا تبين بعد الفحص والدراسة أن وجود البيانات الشاذة ناتج عن خطأ في مرحلة جمع البيانات أو في مرحلة المعالجة فعلى الباحث تصحيح هذه الأخطاء وإعادة حل النموذج. وأما إذا كانت البيانات الشاذة بيانات حقيقية فهناك عدد من الحلول المقترحة نذكر منها ما يلي:

- إعادة توصيف النموذج (Model re-specification) وذلك إما بإضافة أو حذف متغيرات مستقلة.
- إجراء تحويلة/تحويلات إما للمتغير التابع و/أو بعض المتغيرات المستقلة كتحويلة اللوغاريتم والمعكوس (Reciprocal).
- حذف المشاهدات الشاذة إذا كان حجم العينة (عدد المشاهدات) كبيراً وإعادة حل النموذج.
- جمع بيانات إضافية لزيادة حجم العينة وبالتالي تقليل أثر وجود البيانات الشاذة. إلا أن هذا الحل قد يصاحبه بعض المصاعب كارتباط الظاهرة محل الدراسة بفترة زمنية محددة مما يعني الانتظار حتى حلول الظاهرة مرة أخرى وقد تكون هذه الفترة غير معلومة كحدوث الكوارث فضلاً عن التكلفة المضافة لجمع هذه البيانات.
- يقترح لويس-بيك (Lewis-Beck (1993), pp32-33) أن يتم تقديم تقريرين لنتائج نموذج الانحدار أحدهما يحتوي على نتائج النموذج باستخدام كل الحالات والآخر يحتوي على نتائج النموذج بعد استبعاد الحالات الشاذة.
- توجد بعض الطرق المقاومة (Robust) للبيانات الشاذة كطريقة الانحرافات المطلقة الصغري (Method of Least Absolute Deviations). وفي هذه الطريقة يتم الحصول على تقدير معالم نموذج الانحدار بتدنية القيم المطلقة للانحرافات أي:

$$\sum_{i=1}^{n}\left|y_i - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)\right|$$

وبما أنه يتم تدنية الانحرافات المطلقة وليس مربع الانحرافات، كما يتم ذلك في طريقة المربعات الصغرى، نجد أن هذه الطريقة تقلل من أثر المشاهدات الشاذة. ويعاب على هذه الطريقة أنه ربما نفقد خصائص طريقة المربعات الصغرى الحميدة المتمثلة في عدم التحيز، الاتساق والكفاءة.

تمارين

١. باستخدام طرق جبر المصفوفات برهن على أن قيمة الرافعة رقم (i) في نموذج الانحدار الخطي البسيط تأخذ الصيغة التالية:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

٢. البيانات التالية تختص بأعمار ٦ أزواج تم اختيارهم عشوائياً بهدف قياس العلاقة بين عمر الزوجة وعمر الزوج:

عمر الزوج (X_i)	٤٠	٣٠	٦٠	٢٣	١٩	٣٢
عمر الزوجة (Y_i)	٤٠	٢٧	٤٠	٢٥	١٧	٣١

المطلوب:

- تقدير معلمتي نموذج انحدار عمر الزوجة على عمر الزوج:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- هل توجد بيانات شاذة في المتغيرين التابع والمستقل؟
- هل البيانات الشاذة – إن وجدت - مؤثرة؟
- اقترح حلاً مناسباً في حالة وجود بيانات شاذة.

الفصل الخامس

استخدام المتغيرات الصوّرية في تحليل الانحدار الخطي

٥-١ مقدمــــة:

يتطلب تحليل الانحدار الخطي أن تكون المتغيرات المستقلة متغيرات كمّية، إلا أنه في الواقع نجد متغيرات نوعية كثيرة يمكن أن تسهم في تفسير التغير أو التباين في المتغير التابع. فمثلاً المصروفات المعيشية للأسرة تعتمد على متغيرات كمية كالدخل، حجم الأسرة، .. الخ وتعتمد أيضاً على متغيرات نوعية كنوع الحي (حي راقٍ، حي شعبي)، مستوى تعليم رب الأسرة (أمي، متعلم)، ملكية السكن (مالك، مستأجر)، .. الخ. ويتطلب إدخال المتغيرات النوعية في نموذج الانحدار تحويلها إلى متغيرات تعرف بالمتغيرات الصورية (Dummy variables). والمتغير الصوري هو متغير يأخذ عدداً محدداً من القيم تمثل فئات/صفات المتغير النوعي. فمثلاً باستخدام الترميز الثنائي (Binary coding) يأخذ المتغير الصوري القيمة "واحد" في حالة وجود خاصية معينة والقيمة "صفر" عند غياب هذه الخاصية. وتعني كلمة صوري (Dummy) أن القيم التي تأخذها هذه المتغيرات لا تشير إلى قياس حقيقي ذي معنى بل تستخدم فقط لتمييز صفات المتغير النوعي. وبتحويل صفات المتغير النوعي إلى متغيرات صوّرية تأخذ القيم "واحد" أو "صفر" تصبح هذه المتغيرات كَميةً ذات فئات متساوية يمكن استخدامها في تحليل الانحدار.

يعالج هذا الفصل موضوع المتغيرات النوعية من حيث طرق ترميزها لتحويلها إلى متغيرات صورية واستخدامها كمتغيرات مفسرة في نموذج الانحدار الخطي.

٥-٢ طرق الترميز (Coding schemes):

توجد طريقتان لترميز المتغيرات النوعية هما: **الترميز الثنائي**؛ حيث يأخذ المتغير الصوري إما القيمة "واحد" أو القيمة "صفر"، و**الترميز الثلاثي** ويأخذ المتغير الصوري إما القيمة "موجب واحد" أو "سالب واحد" أو "صفر" . وتجدر الإشارة هنا إلى أنه يتم الحصول على نفس النتائج باستخدام أي من هاتين الطريقتين للترميز، إلا أن الطريقة الأولى – الترميز الثنائي- هي الأوسع استخداماً. وفيما يلي نورد أمثلة لتحويل المتغيرات النوعية إلى متغيرات صورية:

- **مثال - حالة التدخين (يدخن ، لا يدخن):**

باستخدام الترميز الثنائي يمكن تحويل متغير حالة التدخين إلى متغير صوري كما يلي:

$$D = \begin{cases} ١ & \text{إذا كان الشخص يدخن.} \\ \text{صفر} & \text{إذا كان الشخص لا يدخن.} \end{cases}$$

وتسمى الفئة التي تأخذ القيمة "صفر" بفئة الأساس (Base category)، ويمكن تحديد أي من الفئتين كفئة أساس. ويعد اختيار فئة الأساس أمراً تحكمياً يعتمد على الباحث. ففي هذا المثال يمكن تحديد الفئة "يدخن" كفئة أساس بدلاً من الفئة "لا يدخن" كما يلي:

$$D = \begin{cases} \text{١} & \text{إذا كان الشخص لا يدخن.} \\ \text{صفر} & \text{إذا كان الشخص يدخن.} \end{cases}$$

وباستخدام الترميز الثلاثي يمكن تحويل متغير حالة التدخين إلى متغير صوري كما يلي:

$$D = \begin{cases} \text{+١} & \text{إذا كان الشخص يدخن.} \\ \text{-١} & \text{إذا كان الشخص لا يدخن.} \end{cases}$$

أو

$$D = \begin{cases} \text{+١} & \text{إذا كان الشخص لا يدخن.} \\ \text{-١} & \text{إذا كان الشخص يدخن.} \end{cases}$$

وفي هذه الحالة تسمى الفئة التي تأخذ القيمة "-١" بفئة الأساس.

- **مثال – متغير الحالة الزواجية:**

يمكن استخدام أكثر من متغير صوري واحد لتمثيل فئات/صفات متغير نوعي واحد ففي هذا المثال يمكن تحويل متغير الحالة الزواجية (متزوج، أعزب، أرمل، مطلق) إلى ثلاث متغيرات صورية كما يلي:

$$D_1 = \begin{cases} \text{١} & \text{إذا كان الشخص متزوجاً.} \\ \text{صفر} & \text{إذا كان الشخص بخلاف ذلك (أعزب، أرمل، مطلق).} \end{cases}$$

و

$$D_2 = \begin{cases} \text{١} & \text{إذا كان الشخص أعزب.} \\ \text{صفر} & \text{إذا كان الشخص بخلاف ذلك (متزوج، أرمل، مطلق).} \end{cases}$$

و

$$D_3 = \begin{cases} ١ & \text{إذا كان الشخص أرملاً.} \\ \text{صفر} & \text{إذا كان الشخص بخلاف ذلك (متزوج، أعزب، مطلق).} \end{cases}$$

كما يمكن تحديد قيم المتغيرات الصورية باستخدام الجدول التالي:

الحالة الزواجية/المتغير الصوري	D_1	D_2	D_3
متزوج	1	0	0
أعزب	0	1	0
أرمل	0	0	1
مطلق	0	0	0

ويلاحظ في هذا المثال أن الحالة الزواجية "مطلق" هي فئة الأساس إذ تأخذ القيمة "صفر" عند كل المتغيرات الصورية الثلاثة. وكما سبق ذكره يمكن أن تكون أي من هذه الفئات فئةً للأساس. وباستخدام الترميز الثلاثي يمكن تعريف المتغيرات الصورية التالية:

$$D_1 = \begin{cases} +١ & \text{إذا كان الشخص متزوجاً.} \\ -١ & \text{إذا كان الشخص مطلقاً.} \\ 0 & \text{إذا كان الشخص بخلاف ذلك.} \end{cases}$$

و

$$D_2 = \begin{cases} +١ & \text{إذا كان الشخص أعزباً.} \\ -١ & \text{إذا كان الشخص مطلقاً.} \\ 0 & \text{إذا كان الشخص بخلاف ذلك.} \end{cases}$$

و

$$D_3 = \begin{cases} +١ & \text{إذا كان الشخص أرملاً.} \\ -١ & \text{إذا كان الشخص مطلقاً.} \\ 0 & \text{إذا كان الشخص بخلاف ذلك.} \end{cases}$$

كما يمكن تعريف المتغيرات باستخدام الجدول التالي:

الحالة الزواجية/المتغير الصوري	D_1	D_2	D_3
متزوج	1	0	0
أعزب	0	1	0
أرمل	0	0	1
مطلق	-1	-1	-1

ويلاحظ من الجدول أن فئة الأساس "مُطلق" تأخذ القيمة "سالب واحد صحيح" عند كل المتغيرات الصورية.

وبصورة عامة إذا كان المتغير النوعي يحتوي على عدد (m) من الفئات أو الصفات فإنه يمكن تحويله إلى عدد (m-1) من المتغيرات الصورية في حالة اشتمال نموذج الانحدار على المعامل الثابت (β_0). وفي هذه الحالة يمكن تعريف (m-1) متغير صوري كما يلي:

الصفة/الفئة	D_1	D_2	D_3	.	.	.	D_{m-1}
1	1	0	0	.	.	.	0
2	0	1	0	.	.	.	0
3	0	0	1	.	.	.	0
.	.	.	.	.			.
.	.	.	.		.		.
.	.	.	.			.	.
m-1	0	0	0	.	.	.	1
m	0	0	0	.	.	.	0

ومن الجدول أعلاه نجد أن الفئة (m) هي فئة الأساس. وأما إذا قمنا بتعريف عدد (m) متغير صوري بعدد صفات المتغير النوعي في حالة اشتمال نموذج الانحدار على المعامل الثابت (Intercept) فإننا نواجه بمشكلة الارتباط الخطي التام([1]) (Perfect Multicollinearity)، المشكلة التي يتعذر بوجودها استخدام طريقة المربعات الصغرى لتقدير معالم النموذج. فمثلاً إذا كان لدينا متغير نوعي ذو صفتين كمتغير الجنس وقمنا بتعريف متغيرين صوريين لتمثيل صفتي المتغير، نجد أن مصفوفة البيانات تأخذ الشكل التالي:

[1] يعالج الفصل السابع مشكلة الارتباط الخطي بشيء من التفصيل.

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ . & . & . \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ . & . & . \\ 1 & 0 & 1 \end{pmatrix}$$

حيث يحتوي العمود الأول على القيمة "واحد" لتقدير المعامل الثابت والعموديين الثاني والثالث يحتويان على قيم المتغيرين الصوريين. ويلاحظ من مصفوفة البيانات أنه يمكن الحصول على قيم العمود الثاني بطرح قيم العمود الثالث من قيم العمود الأول، وكذلك يمكن الحصول على قيم العمود الثالث من مصفوفة البيانات بطرح قيم العمود الثاني من قيم العمود الأول. وبالتالي نجد أن محددة هذه المصفوفة تساوي الصفر ومن ثم لا يمكن إيجاد معكوس المصفوفة $\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}$ وعدم إمكانية استخدام طريقة المربعات الصغرى لتقدير معالم النموذج. كما يجب ملاحظة عدم بروز هذه المشكلة في حالة عدم احتواء نموذج الانحدار على المعامل الثابت (β_0) حيث تأخذ مصفوفة البيانات في هذا المثال الشكل التالي:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ . & . \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ . & . \\ 0 & 1 \end{pmatrix}$$

٥-٣ استخدام المتغيرات الصورية كمتغيرات مستقلة في نموذج الانحدار الخطي:

يمكن بناء نموذج الانحدار الخطي على أساس وجود عدد من المتغيرات المستقلة الكمية والنوعية. سنتناول في هذا الجزء نماذج انحدار تضم متغيرات نوعية فقط، ونماذج انحدار تضم متغيرات كمية ونوعية، ونماذج انحدار تضم متغيرات كمية ونوعية ومتغيرات تفاعل بينهما.

٥-٣-١ نموذج انحدار يحتوي على متغير مستقل نوعي واحد:

من الممكن أن يحتوي نموذج الانحدار على متغير مستقل نوعي واحد فقط. فإذا كان المتغير النوعي ذا فئتين كمتغير النوع (ذكر/أنثى) تأخذ معادلة الانحدار في هذه الحالة الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i \qquad (5\text{-}1)$$

حيث إن:

Y_i = المتغير التابع.

D_i = متغير صوري يأخذ القيمة "١" في حالة وجود الخاصية والقيمة "صفر" في حالة غياب الخاصية.

ε_i = حد الخطأ العشوائي.

ويمكن إيجاد القيمة المتوقعة للمتغير التابع لفئتي المتغير النوعي كما يلي:

- القيمة المتوقعة للمتغير التابع في حالة وجود الخاصية ($D_i = 1$) هي:

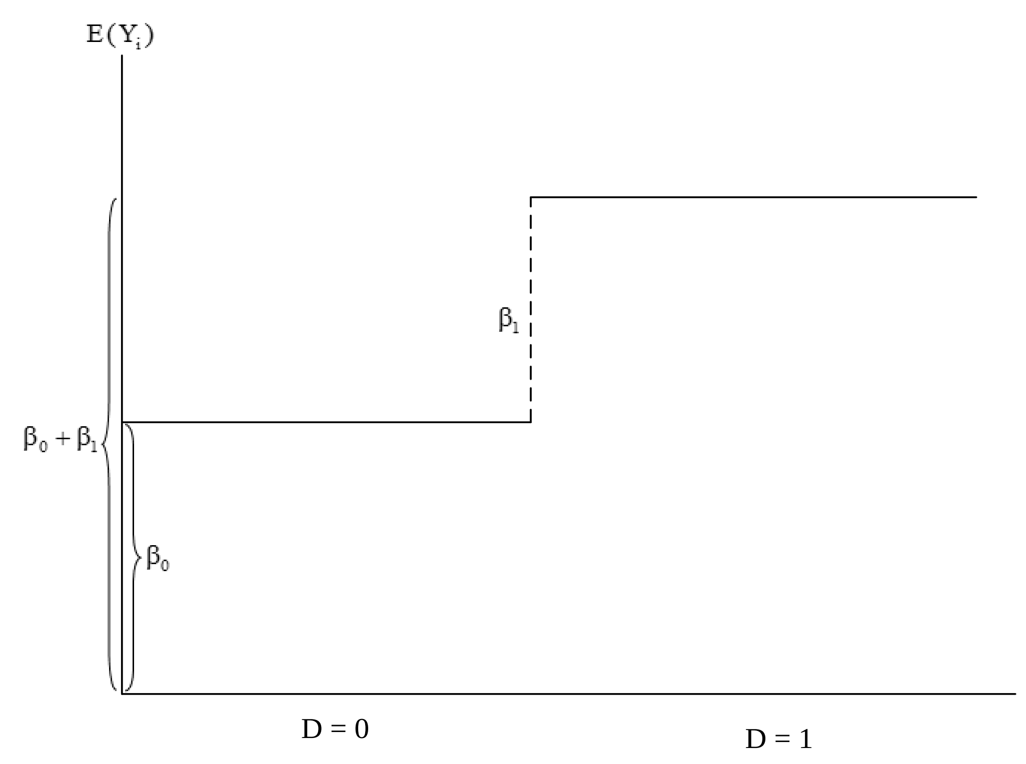
$$E\left(Y_i | D_i = 1\right) = \beta_0 + \beta_1 D_i + \varepsilon_i$$

- والقيمة المتوقعة للمتغير التابع في حالة غياب الخاصية ($D_i = 0$) هي:

$$E\left(Y_i | D_i = 0\right) = \beta_0$$

ويلاحظ أن المعامل الثابت (β_0) يشير إلى القيمة المتوقعة للمتغير التابع في حالة غياب الخاصية (D_i=0) وأن ميل الانحدار (β_1) يشير إلى الفرق بين القيمة المتوقعة للمتغير التابع في حالة غياب الخاصية والقيمة المتوقعة في حالة وجود الخاصية (D_i=1) (شكل رقم ٥-١).

ويتم إجراء اختبار ميل الانحدار كالمعتاد، أي أن يتم اختبار فرض العدم $(H_1 : \beta_1 = 0)$ ضد الفرض البديل $(H_1 : \beta_1 \neq 0)$. فإذا اتضح من الاختبار قبول فرض العدم حكمنا بعدم وجود فرق بين قيمتي المتغير التابع المتوقعة في حالتي غياب ووجود الخاصية، لأن معنى $(\beta_1 = 0)$ أن تكون قيمة $(Y = \beta_0)$ أي أن لـ Y قيمة ثابتة. وأما إذا رفضنا فرض العدم نحكم بأن الفرق بين قيم ا لمتغير التابع المتوقعة تختلف جوهرياً حسب قيمتي المتغير الصوري أو حسب فئتي المتغير النوعي. ويلاحظ أن هذا الاختبار يكافئ اختبار (t) لاختبار الفرق بين متوسطين مجتمعين.



شكل رقم (٥-١): القيم المتوقعة للمتغير التابع حسب قيم المتغير الصوري (D_i) - حالة متغير صوري ذو صفتين.

وفي حالة اشتمال المتغير النوعي على عدد (m) من الصفات (m أكبر من ٢)، فإنه يتم تعريف عدد (m-1) متغير صوري ويأخذ نموذج الانحدار في هذه الحالة الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \ldots + \beta_1 D_{m-1,i} + \varepsilon_i \qquad (5\text{-}2)$$

حيث إن:

Y_i = المتغير التابع.

D_{ji} = 1 في حالة وجود الصفة j ، j=1,2,...,m-1
0 بخلاف ذلك.

ε_i = حد الخطأ العشوائي.

ويمكن حساب القيمة المتوقعة للمتغير التابع لكل صفة من صفات المتغير النوعي على النحو التالي:

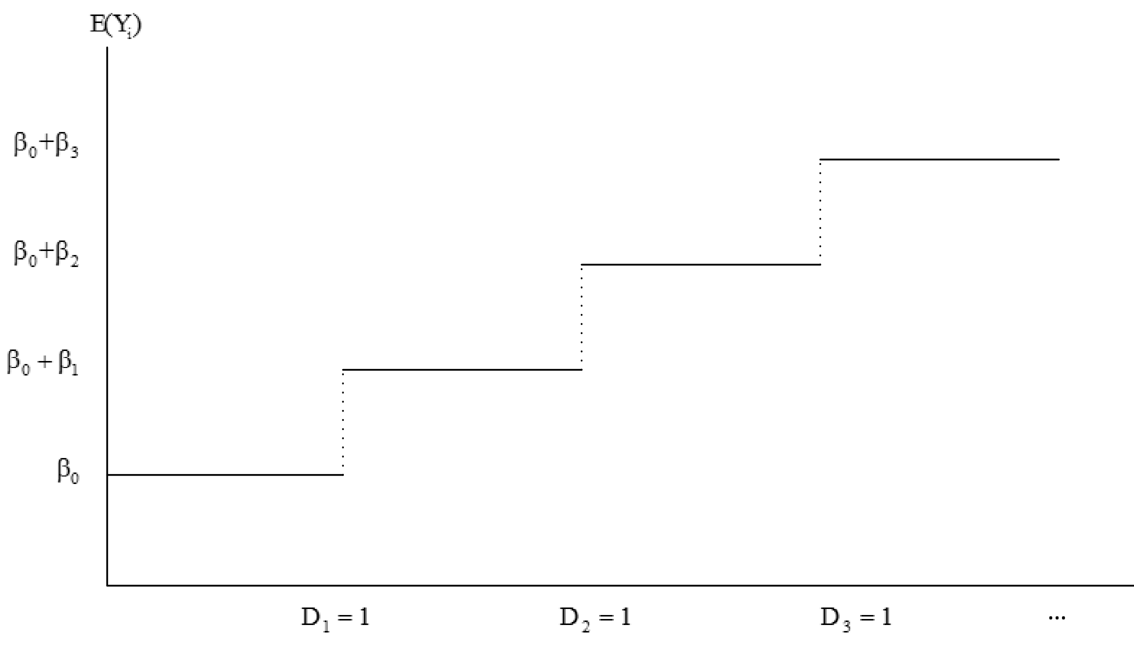
- القيمة المتوقعة للمتغير التابع لفئة الأساس (m) هي:

$$E\left(Y_i|D_1=D_2=\ldots=D_{m-1}=0\right) = \beta_0$$

- والقيمة المتوقعة للمتغير التابع للفئة الأولى هي:

$$E\left(Y_i | D_1=1,\ D_2=D_3=...=D_{m-1}=0\right)=\beta_0+\beta_1$$

وهكذا يمكن إيجاد القيم المتوقعة للمتغير التابع لبقية الصفات. والشكل رقم (٥-٢) يوضح القيم المتوقعة للمتغيرات حسب صفات المتغير النوعي.



شكل رقم (٥-٢): القيم المتوقعة للمتغير التابع حسب المتغيرات الصورية -حالة متغير صوري ذي صفات متعددة.

وبعد بناء النموذج يتم إجراء اختبار دلالة الانحدار ككل وذلك باختبار فرض العدم $\left(H_0:\ \beta_1=\beta_2=\ldots=\beta_{m-1}=0\right)$ في مقابل الفرض البديل "ليس كل قيم معاملات النموذج مساوية للصفر". فإذا أشار الاختيار بقبول فرض العدم حكمنا بعدم وجود علاقة خطية بين المتغير التابع والمتغير النوعي ذي الصفات المتعددة. وأما إذا رفضنا فرض العدم لصالح الفرض البديل "ليس كل قيم معالم النموذج مساوية للصفر" نحكم بأن هناك علاقة خطية بين المتغير التابع والمتغير النوعي، ومن ثم يتم اختبار دلالة معاملات الانحدار الجزئية للمتغيرات الصورية حسب الإحصاءة التالية:

$$\frac{\widehat{\beta}_j}{s.e.\ (\widehat{\beta}_j)} \quad t_{n-m}, \quad j=1,2,...,m\text{-}1$$

ويعرف هذا النموذج بنموذج تحليل التباين في اتجاه واحد (One-way Analysis of Variance).

مثال (٥-١):

في هذا المثال سنقوم بتحليل بيانات افتراضية عن العوامل المؤثرة على مستوى التحصيل الأكاديمي لعينة من تلاميذ الصف السادس من المرحلة الابتدائية. حيث يوضح الجدول رقم (٥-١) بيانات درجات الطلاب من خمس مواد (درجات كل مادة تساوي ١٠٠درجة)، نوع المدرسة (حكومي/خاص)، مستوى تعليم رب الأسرة ومتوسط ساعات الاستذكار اليومي. سيتم بناء نماذج انحدار لقياس تأثير هذه المتغيرات الثلاثة على مستوى تحصيل التلميذ. المطلوب أولاً تحويل المتغيرين النوعيين إلى متغيرات صورية ومن ثم بناء نموذج انحدار خطي لقياس أثر نوع المدرسة على درجات التلاميذ.

الحل:

- تحويل المتغيرات النوعية إلى صورية:

يتم تحويل متغير نوع المدرسة إلى متغير صوري كما يلي:

D = { ١ إذا كانت المدرسة خاصةً.
صفر إذا كانت المدرسة حكومية.

كما يمكن تحويل متغير مستوى تعليم رب الأسرة كما يلي:

مستوى التعليم	متغير- ص١	متغير-ص٢
غير متعلم	1	0
تعليم متوسط (ابتدائي أو مرحلة متوسطة)	0	1
تعليم ثانوي وما فوق	0	0

ويلاحظ من الجدول أعلاه أن فئة الأساس هي مستوى التعليم "تعليم ثانوي وما فوق".

جدول رقم (٥-١): الدرجات المتحصلة، نوع المدرسة، مستوى تعليم رب الأسرة وساعات الاستذكار لعدد (٤٣) طالباً (بيانات افتراضية).

رقم المشاهدة	درجات التلميذ	نوع المدرسة	مستوى تعليم رب الأسرة	متوسط ساعات الاستذكار اليومي
1	249	خاصة	غير متعلم	1
2	208	حكومية	غير متعلم	1
3	336	خاصة	تعليم ثانوى وما فوق	1
4	203	حكومية	غير متعلم	1
5	211	حكومية	غير متعلم	1
6	375	خاصة	تعليم متوسط	1
7	211	حكومية	تعليم متوسط	1
8	189	حكومية	غير متعلم	1
9	219	حكومية	تعليم ثانوى وما فوق	1
10	221	حكومية	تعليم متوسط	1
11	269	خاصة	غير متعلم	1
12	384	خاصة	تعليم متوسط	1
13	356	خاصة	تعليم ثانوى وما فوق	2
14	205	حكومية	غير متعلم	2
15	364	خاصة	تعليم ثانوى وما فوق	2
16	238	حكومية	تعليم ثانوى وما فوق	2
17	231	حكومية	تعليم متوسط	2
18	238	حكومية	تعليم متوسط	2
19	246	حكومية	تعليم ثانوى وما فوق	2
20	287	خاصة	غير متعلم	3
21	222	حكومية	غير متعلم	3
22	384	خاصة	تعليم ثانوى وما فوق	3
23	249	حكومية	تعليم متوسط	3
24	411	خاصة	تعليم ثانوى وما فوق	3
25	305	خاصة	غير متعلم	3
26	266	حكومية	تعليم متوسط	3
27	313	خاصة	غير متعلم	3
28	398	خاصة	تعليم ثانوى وما فوق	3
29	243	حكومية	غير متعلم	3
30	260	حكومية	غير متعلم	3
31	286	حكومية	تعليم متوسط	4
32	416	خاصة	تعليم ثانوى وما فوق	4
33	427	خاصة	تعليم متوسط	4
34	456	خاصة	تعليم ثانوى وما فوق	4
35	267	حكومية	غير متعلم	4
36	303	حكومية	تعليم متوسط	4
37	434	خاصة	تعليم ثانوى وما فوق	4
38	287	حكومية	غير متعلم	4
39	473	خاصة	تعليم متوسط	5
40	322	حكومية	تعليم متوسط	5
41	461	خاصة	تعليم ثانوى وما فوق	5
42	500	خاصة	تعليم ثانوى وما فوق	5
43	338	حكومية	تعليم متوسط	4

جدول رقم (٥-٢): الدرجات المتحصلة، نوع المدرسة، مستوى تعليم رب الأسرة وساعات الاستذكار - تحويل المتغيرات النوعية إلى صورية

رقم المشاهدة	درجات التلميذ	نوع المدرسة	تعليم-ص١	تعليم -ص٢	عدد ساعات الاستذكار
1	249	1	1	0	1
2	208	0	1	0	1
3	336	1	0	0	1
4	203	0	1	0	1
5	211	0	1	0	1
6	375	1	0	1	1
7	211	0	0	1	1
8	189	0	1	0	1
9	219	0	0	0	1
10	221	0	0	1	1
11	269	1	1	0	1
12	384	1	0	1	1
13	356	1	0	0	2
14	205	0	1	0	2
15	364	1	0	0	2
16	238	0	0	0	2
17	231	0	0	1	2
18	238	0	0	1	2
19	246	0	0	0	2
20	287	1	1	0	3
21	222	0	1	0	3
22	384	1	0	0	3
23	249	0	0	1	3
24	411	1	0	0	3
25	305	1	1	0	3
26	266	0	0	1	3
27	313	1	1	0	3
28	398	1	0	0	3
29	243	0	1	0	3
30	260	0	1	0	3
31	286	0	0	1	4
32	416	1	0	0	4
33	427	1	0	1	4
34	456	1	0	0	4
35	267	0	1	0	4
36	303	0	0	1	4
37	434	1	0	0	4
38	287	0	1	0	4
39	473	1	0	1	5
40	322	0	0	1	5
41	461	1	0	0	5
42	500	1	0	0	5
43	338	0	0	1	4

- **نموذج انحدار درجات التلاميذ على نوع المدرسة:**

يوضح الإطار رقم (٥-١) نتائج نموذج انحدار درجات التلاميذ على نوع المدرسة. وتشير النتائج إلى أن متغير نوع المدرسة يؤثر على درجات التلاميذ بمستوى معنوي (p-value = 0.000). ويأخذ النموذج المقدر الصيغة التالية:

$$\hat{y}_i = 246.22 + 133.68\ D_i$$
$$(0.000) \quad (0.000)$$

حيث إن: $\hat{y}_i$ = درجات التلميذ المقدرة و D_i = متغير صوري يمثل متغير نوع المدرسة (١=مدرسة خاصة، ٠= مدرسة حكومية). وتشير الأرقام داخل الأقواس تحت معاملي الانحدار إلى قيم الاحتمال (p-values).

وباستخدام هذه المعادلة نجد أن:

القيمة المتوقعة لدرجات التلميذ في المدرسة الحكومية (D=0) هي:

$$\hat{y}_i = 246.22 + 133.68 \times 0 = 246.22$$

والقيمة المتوقعة لدرجات التلميذ في المدرسة الخاصة (D=0) هي:

$$\hat{y}_i = 246.22 + 133.68 \times 1 = 379.9$$

ومما سبق يلاحظ أن المعامل الثابت يشير إلى درجات التلميذ في المدرسة الحكومية. أما معامل الانحدار (معامل المتغير الصوري) فيشير إلى الفرق بين متوسط درجات التلميذ في المدرسة الحكومية والمدرسة الخاصة، أي أن درجات التلميذ في المدرسة الخاصة تزيد في المتوسط بنحو ١٣٣,٧ درجة عن درجات التلميذ في المدرسة الحكومية (انظر الشكل ٥-٣).

إطار رقم (٥-١): نتائج نموذج انحدار درجات التلميذ على نوع المدرسة؛ (مخرجات برنامج إكسل).

Regression Statistics	
Multiple R	0.772792692
R Square	0.597208545
Adjusted R Square	0.587384363
Standard Error	56.07958479
Observations	43

ANOVA

	df	SS	MS	F	Significance F
Regression	1	191178.566	191178.6	60.78965	0.0000
Residual	41	128941.713	3144.92		
Total	42	320120.2791			

	Coefficients	Standard Error	t Stat	P-value
Intercept	246.2173913	11.69340176	21.0561	0.0000
نوع المدرسة	133.6826087	17.1458927	7.796772	0.0000

وفي حالة تغيير فئة الأساس لتكون "مدرسة خاصة" بدلاً من "مدرسة حكومية" كما يلي:

$D = \begin{cases} ١ & \text{إذا كانت المدرسة حكومية.} \\ \text{صفر} & \text{إذا كانت المدرسة خاصة.} \end{cases}$

فإنه يتم الحصول على نموذج الانحدار المقدر التالي:

$$\hat{y}_i = 379.9 - 133.68\, D_i$$
$$(0.000) \quad (0.000)$$

حيث إن: $\hat{y}_i$ = درجات التلميذ المقدرة و D_i = متغير صوري يمثل متغير نوع المدرسة (١=مدرسة حكومية، ٠= مدرسة خاصة)، والأرقام داخل الأقواس تحت معاملي الانحدار هي قيم الاحتمال (P-values).

ويلاحظ أن إشارة معامل الانحدار قد تغيرت وأصبحت سالبة وأن قيمة المعامل الثابت زادت بـ ١٣٣,٦٨ (قيمة معامل الانحدار). وباستخدام هذه المعادلة يتم الحصول على نفس القيم المتوقعة لدرجات التلاميذ في المدرسة الحكومية والخاصة كما يلي:

القيمة المتوقعة لدرجات التلميذ في المدرسة الحكومية (D=1) هي:

$$\hat{y}_i = 379.9 - 133.68 \times 1 = 246.22$$

والقيمة المتوقعة لدرجات التلميذ في المدرسة الخاصة (D=0) هي:

$$\hat{y}_i = 379.9 - 133.68 \times 0 = 379.9$$

٥-٣-٢ نموذج انحدار يشتمل على متغير كمي واحد ومتغير نوعي ذي صفتين:

يأخذ نموذج الانحدار الذي يشتمل على متغير كمي واحد ومتغير نوعي بصفتين الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \varepsilon_i \qquad (5-3)$$

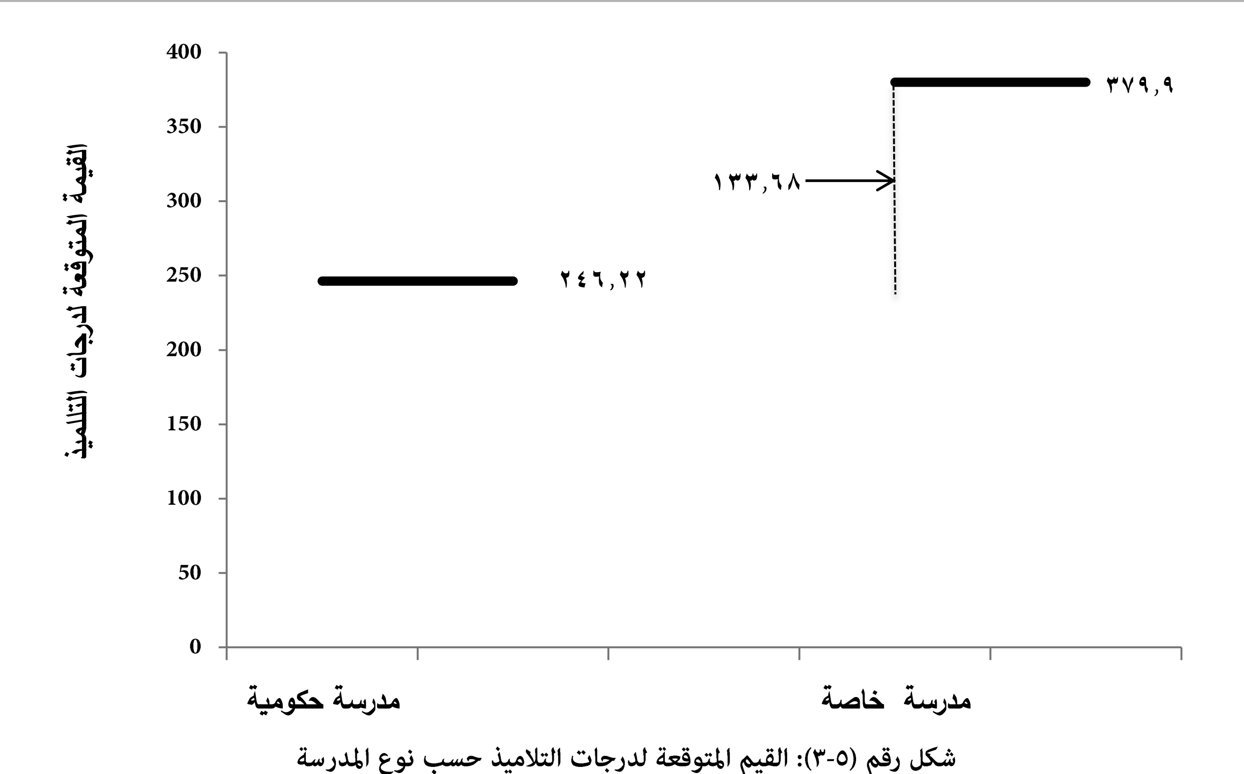
حيث إن:

Y_i = المتغير التابع.

X_i = متغير كمي.

D_i = متغير صوري يأخذ القيمة "١" في حالة وجود الصفة الأولى والقيمة "صفر" في حالة وجود الصفة الثانية.

ε_i = حد الخطأ العشوائي.



شكل رقم (٥-٣): القيم المتوقعة لدرجات التلاميذ حسب نوع المدرسة

ويتم حساب القيمة المتوقعة للمتغير التابع لأي من الصفتين على النحو التالي:

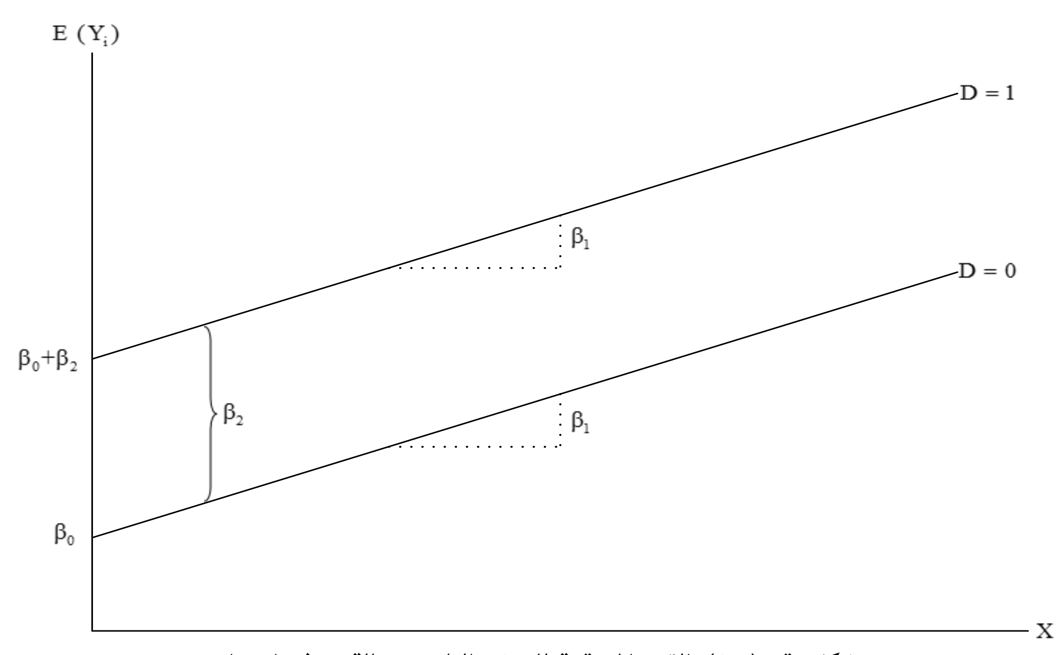
- القيمة المتوقعة للمتغير التابع للصفة الأولى (D=1) هي:

$$E(Y_i|X_i, D_i=1)=\beta_0+\beta_1X_i+\beta_2=(\beta_0+\beta_2)+\beta_1X_i$$

- والقيمة المتوقعة للمتغير التابع للصفة الثانية (D=0) هي:

$$E(Y_i|X_i, D_i=0)=\beta_0+\beta_1X_i$$

ويلاحظ أن إدخال المتغير الصوري (D) في نموذج الانحدار يماثل إجراء نموذجي انحدار أحدهما للصفة الأولى والآخر للصفة الثانية. ويوضح الشكل رقم (٥-٤) أن معامل المتغير الصوري (β_2) يمثل الفاصل الثابت بين خطي الانحدار المتوازيين. وجدير بالذكر أنه يمكن الحصول على نفس النتائج إذا تم تغيير فئة الأساس مع ملاحظة أن إشارة المعامل (β_2) ستتغير؛ فإذا كانت سالبة تصبح موجبة والعكس صحيح مع ثبات القيمة المطلقة للمعامل.



شكل رقم (٥-٤): القيم المتوقعة للمتغير التابع - حالة نموذج انحدار خطي يشتمل على متغيرنوعي ذي صفتين ومتغير كمي واحد.

ولتحديد ما إذا كان المتغير النوعي يؤثر على المتغير التابع بمستوى معنوي يتم اختبار الفرضية التالية:

فرض العدم $H_1 : \beta_2 = 0$ مقابل **الفرض البديل** $H_1 : \beta_2 \neq 0$

ولإجراء هذا الاختبار تستخدم إحصاء **T** حيث

$$|T| = \frac{\hat{\beta}_2}{s.e(\hat{\beta}_2)} \sim t_{n-3}$$

فإذا كانت قيمة T المطلقة أكبر من قيمة توزيع **t** عند درجات حرية (n-3) ومستوى معنوية $(\alpha/2)$ نرفض فرض العدم ونحكم بأن نقطتي تقاطع خطي الانحدار (intercepts) تختلفان اختلافا ذا دلالة إحصائية. وأما إذا كانت القيمة المطلقة لـ T أقل قيمة من توزيع **t** عند درجات حرية (n-3) نحكم بأن نقطتي التقاطع لا تختلفان وبالتالي يكون لدينا خط انحدار واحد.

ويمكن تطوير المعادلة (5.3) ليضم نموذج الانحدار أكثر من متغير كمي واحد ومتغير نوعي بصفتين على النحو التالي:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \beta_{p+1} D_i + \varepsilon_i \qquad (5\text{-}4)$$

حيث إن:

Y_i = المتغير التابع.

$X_1, X_2, \ldots, X_p$ = عدد p متغير كمي.

D_i = متغير صوري يأخذ القيمة "١" في حالة وجود الصفة الأولى و"صفر" في حالة وجود الصفة الثانية.

ε_i حد الخطأ العشوائي.

وفي هذه الحالة يتم إيجاد القيمة المتوقعة للمتغير التابع للصفتين على النحو التالي:

$$E\left(Y_i \mid X_1, X_2, \ldots, X_p, D_i = 1\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \beta_{p+1}$$

و

$$E\left(Y_i \mid X_1, X_2, \ldots, X_p, D_i = 0\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

كما يتم إجراء اختبار معنوية المتغير الصوري (D) بنفس الطريقة التي سبق شرحها.

مثال (٥-٢):

من بيانات المثال السابق أجرِ نموذج انحدار درجات التلميذ على متغيري نوع المدرسة وساعات الاستذكار وفسر النتائج التي تحصل عليها؟

الحل:

يوضح الإطار رقم (٥-٢) نتائج نموذج انحدار درجات التلاميذ على نوع المدرسة (متغير صوري) وساعات الاستذكار (متغير كمي). وتشير النتائج إلى وجود علاقة خطية بين درجات التلميذ وكل من نوع المدرسة وعدد ساعات الاستذكار ذات دلالة إحصائية (p-value =0.000). كما تشير قيم الاحتمال (p-value) المناظرة لمعاملي الانحدار أن كلاً من المتغيرين – نوع المدرسة وعدد ساعات الاستذكار- يسهمان إسهاما جوهرياً في تفسير درجات التلاميذ والتنبؤ بها (p-value =0.000). حيث يفسر هذان المتغيران (٨٤,٤%) من التغير في قيم درجات التلاميذ. وتأخذ معادلة الانحدار الموفقة الصيغة التالية:

$$\hat{y}_i = 164.92 + 119.85 D_i + 32.8 x_i$$

حيث إن: $\hat{y}_i$ = درجات التلميذ المقدرة وD_i = متغير صوري يمثل نوع المدرسة (١=مدرسة خاصة، ٠=مدرسة حكومية) وx_i = عدد ساعات الاستذكار.

وباستخدام هذه المعادلة يمكن حساب القيم المتوقعة للمتغير التابع كما يلي:

- القيمة المتوقعة لدرجات التلميذ الذي يدرس في مدرسة خاصة هي:

$$E(y_i \mid x_i,\ D=1) = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i$$
$$= 164.92 + 119.85 + 32.8x_i$$
$$= 284.77 + 32.80x_i$$

القيمة المتوقعة لدرجات التلميذ الذي يدرس في مدرسة حكومية هي:

$$E(y_i \mid x_i,\ D=0) = \hat{\beta}_0 + \hat{\beta}_2 x_i$$
$$= 164.92 + 32.8x_i$$

ويلاحظ أن القيمة المتوقعة للمتغير التابع لنوعي المدرسة تتغير تبعاً لعدد ساعات الاستذكار. كما يلاحظ أن ساعات الاستذكار تؤثر على درجات التلميذ في المدرسة الخاصة بنفس المستوى الذي تؤثر به على درجات التلميذ في المدرسة الحكومية ذلك لتساوى قيمتي ميلي الانحدار (=٣٢,٨٠) كما يُلاحظ ذلك من معادلتي القيم المتوقعة أعلاه (انظر الشكل ٥-٥). وكذلك يلاحظ أن متوسط درجات التلاميذ في المدرسة الخاصة يفوق متوسط درجات التلاميذ في المدارس الحكومية بمقدار ثابت$\left(\hat{\beta}_1 = 119.85\right)$. أي أن معادلة الانحدار الموفقة للمدارس الخاصة تختلف عن معادلة انحدار المدارس الحكومية في نقطتي التقاطع فقط؛ حيث تزيد قيمة الأولى عن الثانية بمقدار (١١٩,٨٥).

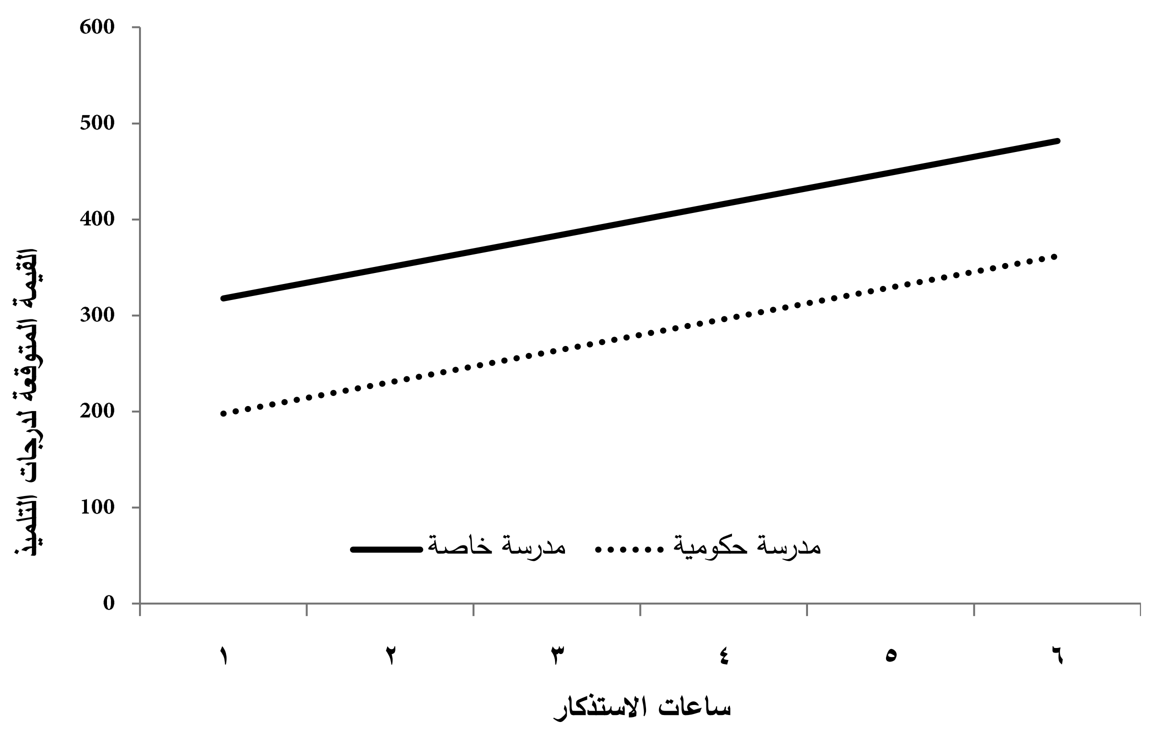
إطار رقم (٥-٢): نتائج نموذج انحدار درجات التلميذ على نوع المدرسة (متغير نوعي) ومتوسط ساعات الاستذكار اليومي؛ (مخرجات برنامج إكسل).

Regression Statistics	
Multiple R	0.918923748
R Square	0.844420855
Adjusted R Square	0.836641898
Standard Error	35.28598846
Observations	43

ANOVA

	Df	SS	MS	F	Significance F
Regression	2	270316.2398	135158.1199	108.551934	0.0000
Residual	40	49804.03926	1245.100981		
Total	42	320120.2791			

	Coefficients	Standard Error	t Stat	P-value
Intercept	164.919475	12.57466034	13.11522303	0.0000
نوع المدرسة	119.8477001	10.92709074	10.96794224	0.0000
عدد ساعات الاستذكار	32.80442237	4.114746103	7.972404992	0.0000



شكل رقم (٥-٥): أثر متغيري نوع المدرسة وساعات الاستذكار على درجات التلاميذ

٥-٣-٣ نموذج انحدار يشتمل على متغير كمي واحد ومتغير نوعي ذي صفات متعددة:

من الممكن أن يضم نموذج الانحدار متغيراً كمياً ومتغيراً نوعياً بصفات متعددة. وفي هذه الحالة يتم أولاً تحديد المتغيرات الصورية والتي يساوي عددها عدد صفات المتغير ناقصاً واحد. ويأخذ نموذج الانحدار في هذه الحالة الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \ldots + \gamma_{m-1} D_{m-1,i} + \varepsilon_i \qquad (5-5)$$

ويتم حساب قيم المتغير التابع المتوقعة حسب قيم المتغيرات الصورية على النحو التالي:

$$E(Y_i|X_i, D_1=D_2=\ldots=D_{m-1}=0) = \beta_0 + \beta_1 X_i$$

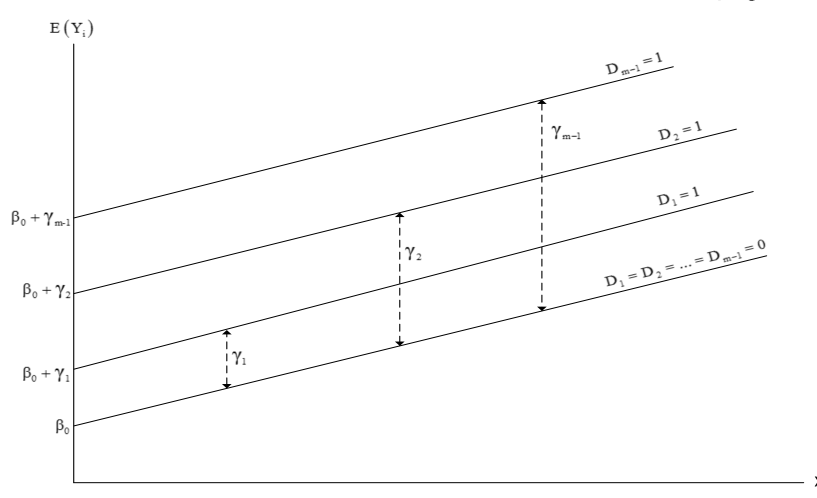
$$E(Y_i|X_i, D_1=1, D_2=D_3=\ldots=D_{m-1}=0) = \beta_0 + \gamma_1 + \beta_1 X_i$$

. .

. .

$$E(Y_i|X_i, D_{m-1}=1, D_1=D_2=\ldots=D_{m-2}=0) = \beta_0 + \gamma_{m-1} + \beta_1 X_i$$

وتوضح المعادلة (5.5) والشكل (٥-٦) أن المعامل الثابت (β_0) يمثل نقطة تقاطع خط انحدار فئة الأساس والمعامل (γ_1) يمثل الفرق في نقطة التقاطع بين خط انحدار فئة الأساس والفئة الأولى والمعامل (γ_2) يمثل الفرق بين نقطتي تقاطع الفئة الثانية وفئة الأساس، وهكذا.



شكل رقم (٥-٦): القيم المتوقعة للمتغير التابع - حالة نموذج يضم متغيراً كمياً ومتغيراً نوعياً بصفات متعددة

ولاختبار معنوية المتغير النوعي يستخدم أسلوب اختبار F الجزئي لاختبار فرض العدم القائل بأن جميع معلمات المتغيرات الصورية مساوية للصفر. ولإجراء هذا الاختبار تتبع الخطوات التالية:

- يتم أولاً بناء نموذجين؛ الأول (النموذج الكامل) ويضم كل المتغيرات الكمية والنوعية والثاني (النموذج المخفض) ويحتوي على المتغير الكمي فقط، أي:

النموذج الكامل هو:

$$Y_i = \beta_0 + \beta_1 X_i + \gamma_1 D_{1i} + \gamma_2 D_{2i} + ... + \gamma_{m-1} D_{m-1,i} + \varepsilon_i \quad (5\text{-}6)$$

والنموذج المخفض هو:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (5\text{-}7)$$

- يتم حساب مجموع مربعات البواقي (RSS) والانحدار (ESS) لكلا النموذجين ومن ثم يتم حساب قيمة إحصاء F0 حيث

$$F_0 = \frac{[ESS(X, D_1, D_2, ..., D_{m-1}) - ESS(X)]/(m-1)}{RSS(X, D_1, D_2, ..., D_{m-1})/(n-m)} \sim F_{m-1,n-m}$$

- تتم مقارنة قيمة F_0 مع قيمة توزيع F بدرجتي حرية (m-1) و(n-m) وعند مستوى معنوية (α). فإذا كانت قيمة F_0 أكبر من قيمة توزيع F بدرجتي حرية (m-1) و(n-m) نرفض فرض العدم وقبول الفرض البديل، أي ليس

كل قيم معلمات المتغيرات الصورية مساوية للصفر؛ وأما إذا كانت قيمة F_0 أقل من قيمة توزيع F نقبل فرض العدم القائل بأن قيم المعلمات $(\gamma_1,\gamma_2,...,\gamma_{m-1})$ مساوية للصفر ونستنتج أن المتغير النوعي لا يسهم في تفسير التغير في المتغير التابع.

مثال (٥-٣):

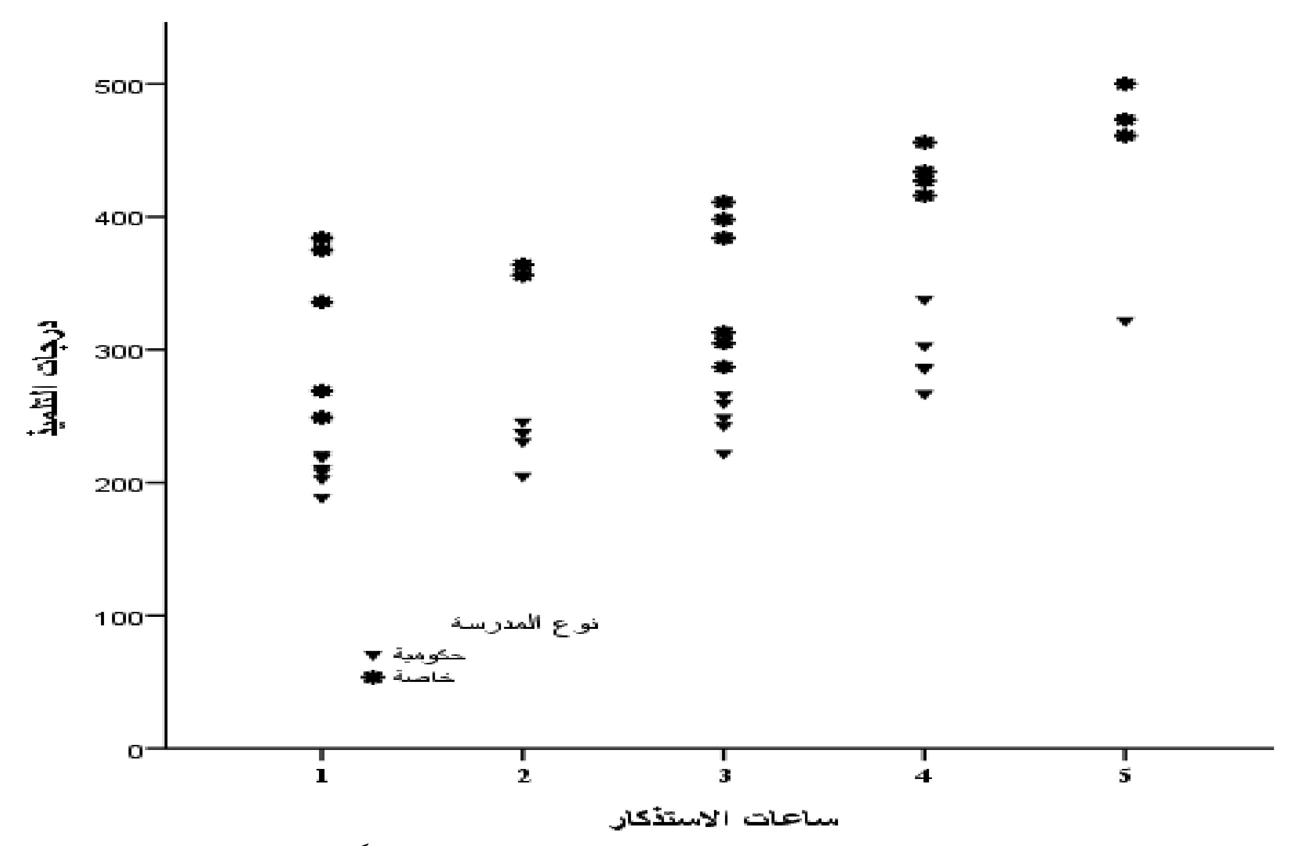
من بيانات المثال السابق أجرِ نموذج انحدار درجات التلميذ على ساعات الاستذكار ومستوى تعليم رب الأسرة وفسر النتائج التي تحصل عليها؟

الحل:

يوضح الشكل رقم (٥-٧) رسم انتشار درجات التلاميذ مع ساعات الاستذكار وفقاً لنوع المدرسة. يوضح الإطار رقم (٥-٣) نتائج نموذج انحدار درجات التلاميذ على ساعات الاستذكار (متغير كمي) ومستوى تعليم رب الأسرة (متغير نوعي ذو ثلاث صفات). ولقد تم في المثال (٥-١) تحويل متغير مستوى تعليم رب الأسرة إلى متغيرين صوريين هما: تعليم -ص١ وتعليم -ص٢. وتشير النتائج المستعرضة بالإطار إلى وجود علاقة خطية بين درجات التلميذ وكل من عدد ساعات الاستذكار ومستوى تعليم رب الأسرة ذات دلالة إحصائية (p-value =0.000). وكما تشير قيم الاحتمال (p-value) المناظرة لمعاملات الانحدار أن كلاً من المتغيرين -عدد ساعات الاستذكار ومستوى تعليم رب الأسرة- يسهمان إسهاما جوهرياً في تفسير درجات التلاميذ والتنبؤ بها (p-value =0.000). ويفسر هذان المتغيران (٦٠,٧%) من التغير في قيم درجات التلاميذ. وتأخذ معادلة الانحدار الموفقة الصيغة التالية:

$$\hat{y}_i = 273.97 + 33.74\, x_i - 102.58 D_{1i} - 61.52 D_{2i}$$

حيث إن: $\hat{y}_i$ = درجات التلميذ المقدرة، x_i = عدد ساعات الاستذكار، D_{1i} = تعليم -ص١ (١= إذا كان رب الأسرة غير متعلم و٠= إذا كان مستوى تعليمه بخلاف ذلك) و D_{2i} = تعليم -ص2 (١= إذا كان تعليم رب الأسرة تعليم متوسط و٠= إذا كان مستوى تعليمه بخلاف ذلك).



شكل رقم (٥-٧): رسم انتشار درجات التلاميذ مع ساعات الاستذكار وفقاً لنوع المدرسة

وباستخدام هذه المعادلة يمكن حساب القيم المتوقعة للمتغير التابع كما يلي:

- القيمة المتوقعة لدرجات التلميذ الذي مستوى تعليم والده ثانوي فما فوق هي:

$$E\left(\hat{y}_i|x_i,\, D_1=D_2=0\right) = \hat{\beta}_0+\hat{\beta}_1 x_i = 273.97 + 33.74x_i$$

- القيمة المتوقعة لدرجات التلميذ الذي مستوى تعليم والده تعليم متوسط هي:

$$E\left(\hat{y}_i|x_i,\, D_1=0,\, D_2=1\right) = \hat{\beta}_0+\hat{\beta}_1 x_i+\hat{\beta}_3 = 273.97 + 33.74x_i - 61.52 = 212.45 + 33.74x_i$$

- القيمة المتوقعة لدرجات التلميذ الذي والده غير متعلم هي:

$$E\left(\hat{y}_i|x_i, D_1=0, D_2=1\right) = \hat{\beta}_0+\hat{\beta}_1X_i+\hat{\beta}_2$$
$$= 273.97 + 33.74x_i - 102.58$$
$$= 171.39 + 33.74x_i$$

إطار رقم (٥-٣): نتائج نموذج انحدار درجات التلميذ على ساعات الاستذكار اليومي ومستوى تعليم رب الأسرة؛ (مخرجات برنامج إكسل).

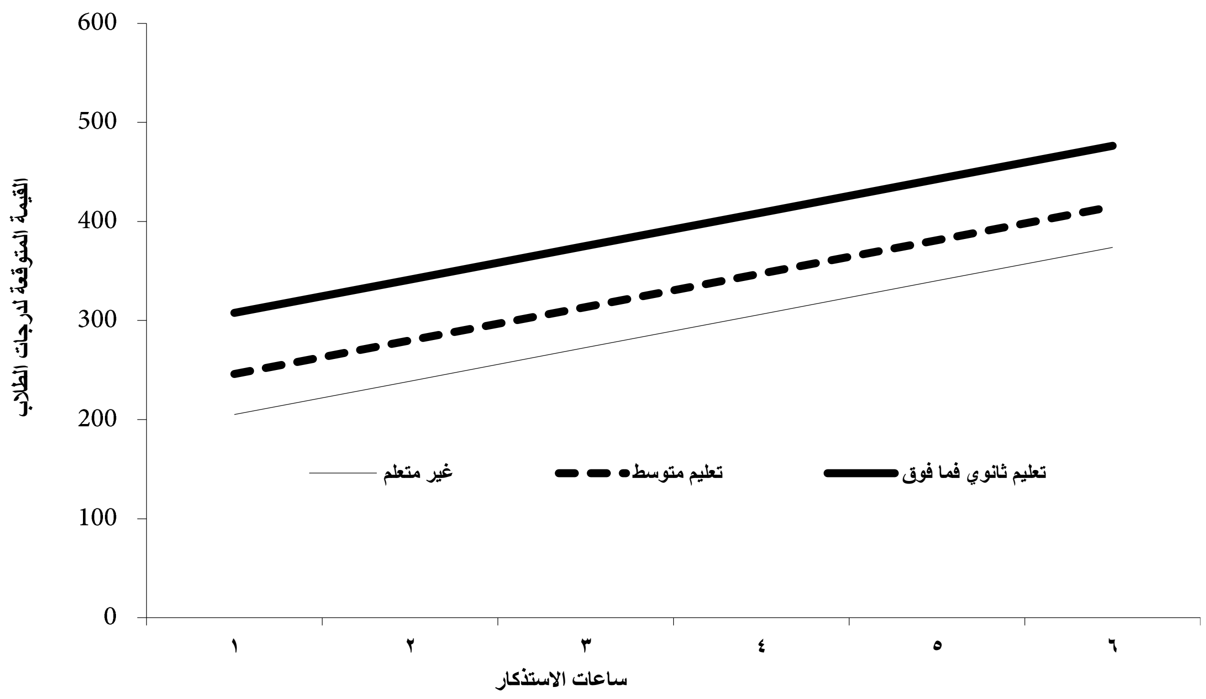
Regression Statistics	
Multiple R	0.77945574
R Square	0.607551251
Adjusted R Square	0.577362886
Standard Error	56.75652071
Observations	43

ANOVA

	df	SS	MS	F	Significance F
Regression	3	194489.476	64829.82533	20.12534447	0.0000
Residual	39	125630.8031	3221.302643		
Total	42	320120.2791			

	Coefficients	Standard Error	t Stat	P-value
Intercept	273.9735594	24.82099406	11.03797691	0.0000
عدد ساعات الاستذكار	33.74073581	6.708593459	5.029479878	0.0000
تعليم-ص١	-102.5858939	21.55375028	-4.759538018	0.0000
تعليم -ص٢	-61.51851887	21.4572997	-2.867020535	0.0066

ويلاحظ أن القيمة المتوقعة للمتغير التابع حسب مستوى تعليم رب الأسرة تتغير تبعاً لعدد ساعات الاستذكار، ذلك لتساوي قيم ميول الانحدار (=٣٣,٧٤) حسب مستويات التعليم (انظر الشكل رقم ٥-٨). كما يلاحظ أن متوسط درجات التلاميذ الذين لدى آبائهم تعليم ثانوي أو فما فوق يفوق متوسط درجات التلاميذ الذين لدى آبائهم تعليم متوسط أو غير متعلمين بـ ٦١,٥٢ و ١٠٢,٥٨ درجة على التوالي. وكذلك نجد أن متوسط درجات التلاميذ الذين لآبائهم تعليم "متوسط" يفوق متوسط درجات التلاميذ الذين لم ينل آباؤهم تعليماً بـ ٤١,٠٦ درجة.



شكل رقم (٥-٨): أثر متغيري مستوى تعليم رب الأسرة وساعات الاستذكار على درجات التلاميذ

٥-٣-٤ التفاعل بين المتغيرات النوعية والكمية:

لقد أوضحنا فيما سبق كيفية تأثير المتغير النوعي على قيم المعامل الثابت أو المعلمة التقاطعية ولكن لم ندرس أثر المتغير النوعي على ميل الانحدار. ولقياس أثر المتغير النوعي على الميل يتم عادة إضافة متغير مستقل مركب عبارة عن مضروب المتغير الصوري في المتغير الكمي ويعرف هذا المتغير بمتغير التفاعل (Interaction regressor). حيث يقيس هذا المتغير الأثر المشترك للمتغيرين النوعي والكمي على المتغير التابع وبصورة عامة يقال لمتغيرين مستقلين أنهما يتفاعلان عندما يكون الأثر الجزئي لأحدهما يعتمد على قيمة الآخر.

ولتوضيح أثر المتغير النوعي على ميل دالة الانحدار سنقوم بدراسة حالة متغير كمي واحد ومتغير نوعي واحد بصفتين. وفي هذه الحالة يأخذ نموذج الانحدار الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + \varepsilon_i \qquad (5\text{-}8)$$

حيث إن:

$\mathbf{Y_i}$ = المتغير التابع.

X_i = متغير كمي.

D_i = متغير صوري يأخذ القيمة "١" إذا كانت المشاهدة تنتمي للصفة الأولى والقيمة "صفر" إذا كانت المشاهدة تنتمي إلى الصفة الثانية.

$D_i X_i$ = متغير مركب يقيس التأثير المشترك للمتغيرين الصوري والكمي على المتغير التابع.

ε_i = حد الخطأ العشوائي.

ويتم إيجاد القيم المتوقعة للمتغير التابع حسب صفتي المتغير النوعي على النحو التالي:

$$E\left(Y_i | X_i, D_i=0\right) = \beta_0 + \beta_1 X_i$$
$$E\left(Y_i / X_i, D_i=1\right) = \beta_0 + \beta_1 X_i + \beta_2 + \beta_3 X_i$$
$$= \left(\beta_0 + \beta_2\right) + \left(\beta_1 + \beta_3\right) X_i$$

ويلاحظ من المعادلة (5.8) أن المعامل (β_2) يقيس الاختلاف في نقطتي التقاطع (المعامل الثابت) لصفتي المتغير النوعي في حين يقيس المعامل (β_3) – معامل متغير التفاعل- الاختلاف في ميلي دالتي الانحدار حسب صفتي المتغير النوعي. وتستخدم هذه المعادلة للإجابة عن الأسئلة التالية:

هل يختلف ميلا دالتي الانحدار لصفتي المتغير النوعي أم أنهما متساويان؟

هل تختلف قيم نقطتي التقاطع (Intercepts) لدالتي الانحدار؟

هل تتطابق دالتا الانحدار لصفتي المتغير النوعي، أي هل لديهما تقاطع وميل واحد؟

ويمكن تمثيل الحالات التي تناظر هذه الأسئلة بالرسوم البيانية الموضحة بالأشكال رقم (٥-٩-أ)، (٥-٩-ب)، (٥-٩-ج) و (٥-٩-د).

وللإجابة عن هذه الأسئلة أعلاه يتم إجراء الاختبارات التالية:

اختبار تساوي ميلي دالتي الانحدار:

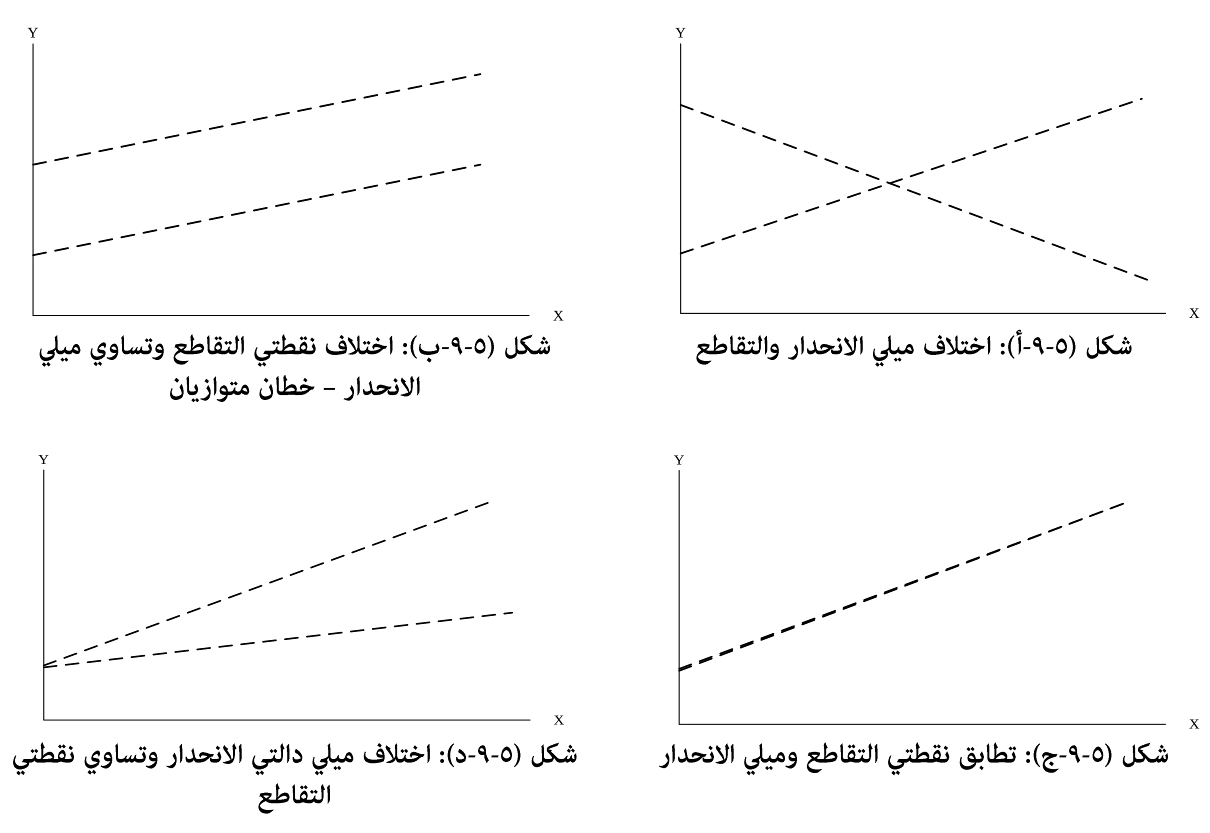
هل هناك اختلاف معنوي بين ميلي دالتي الانحدار لصفتي المتغير النوعي؟ للإجابة عن هذا السؤال يتم إجراء اختبار الفرض التالي:

فرض العدم $\left(H_0 : \beta_3 = 0\right)$ في مقابل الفرض البديل $\left(H_0 : \beta_3 \neq 0\right)$

وفي هذه الحالة تستخدم إحصاء F الجزئي أو اختبار t ، وفي حالة استخدام إحصاء F الجزئي يتم بناء نموذجين أحدهما كامل يضم كل المتغيرات (الكمي، الصوري والتفاعل) والآخر مخفض ويضم المتغير الكمي والصوري فقط. ولإجراء الاختبار يستخدم إحصاء F_0 حيث

$$F_0 = \frac{ESS(X, D, DX) - ESS(X, D)}{RSS(X, D, DX)/(n-3-1)}$$

فإذا كانت قيمة F_0 أكبر من قيمة توزيع F عند درجتي 1 و (n-4) ومستوى معنوية معين، نرفض فرض العدم القائل بأن $(\beta_3 = 0)$ وبذلك نخلص إلى أن ميلي دالتي الانحدار مختلفان. أما إذا كانت قيمة F_0 أقل من قيمة توزيع F الجدولية بدرجتي 1 و (n-4) نحكم بأن ميلي دالتي الانحدار متساويان، أي أن خطي الانحدار لصفتي المتغير النوعي متوازيان.



شكل (٥-٩-أ): اختلاف ميلي الانحدار والتقاطع

شكل (٥-٩-ب): اختلاف نقطتي التقاطع وتساوي ميلي الانحدار - خطان متوازيان

شكل (٥-٩-ج): تطابق نقطتي التقاطع وميلي الانحدار

شكل (٥-٩-د): اختلاف ميلي دالتي الانحدار وتساوي نقطتي التقاطع

شكل (٥-٩): أربع حالات ممكنة لخطي انحدار نموذج يشتمل على متغير كمي ومتغير نوعي ذي صفتين ومتغير تفاعل بينهما.

وفي حالة استخدام اختبار t تستخدم الإحصاءة T حيث

$$T = \frac{\hat{\beta}_3}{s.e.(\hat{\beta}_3)} \sim t_{n-4}$$

فإذا كانت قيمة T المطلقة أكبر من قيمة t الجدولية بدرجة حرية (n-4) وعند مستوى معنوية محدد $(\alpha/2)$ نحكم بأن ميلي خطي الانحدار مختلفان وأما إذا كانت قيمة T المطلقة أقل من القيمة الجدولية نحكم بتوازي خطي الانحدار.

- **اختبار تساوي نقطتي التقاطع (المعامل الثابت):**

يتم اختبار تساوي أو اختلاف نقطتي التقاطع لدالتي الانحدار لصفتي المتغير النوعي باختبار الفرض التالي:

فرض العدم: $(H_1 : \beta_2 = 0)$ **في مقابل** الفرض البديل: $(H_1 : \beta_2 \neq 0)$

وتستخدم في هذه الحالة إما إحصاء F الجزئي أو اختبار t كما سبق شرحهما في حالة اختبار تساوي ميلي الانحدار.

- **اختبار تطابق خطي الانحدار:**

هل خطا الانحدار لصفتي المتغير النوعي متطابقان أم لا؟ للإجابة عن هذا السؤال يتم إجراء اختبار الفرض التالي:

فرض العدم: $(H_0: \beta_2=\beta_3=0)$ **في مقابل** الفرض البديل: ليست كلتا قيمتي المعلمتين β_2 و β_3 مساوية للصفر

وتستخدم في هذه الحالة اختبار إحصاء F الجزئي المتعدد وذلك ببناء نموذج انحدار كامل يضم المتغير الكمي، النوعي ومتغير التفاعل بينهما ونموذج انحدار مخفض يضم المتغير الكمي فقط، أي:

النموذج الكامل:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + \varepsilon_i$$

والنموذج المخفض:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

ويتم حساب مجموع مربعات البواقي والانحدار لكل نموذج ومن ثم يتم حساب إحصاء F_0 حيث

$$F_0 = \frac{[ESS(X, D, DX) - ESS(X)]/2}{RSS(X, D, DX)/(n-4)}$$

فإذا كانت قيمة F_0 أكبر من قيمة F بدرجتي حرية 2 و (n-4) وعند مستوى معنوية معين، نرفض فرض العدم ونحكم بعدم تطابق خطي الانحدار لصفتي المتغير النوعي، أما إذا كانت قيمة F_0 أقل من قيمة F الجدولية نحكم بتطابق خطي الانحدار، وفي هذه الحالة يستخدم النموذج التالي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

مثال (٥-٤):

من بيانات المثال السابق أجرِ نموذج انحدار درجات التلميذ على ساعات الاستذكار ونوع المدرسة ومتغير التفاعل بين نوع المدرسة وساعات الاستذكار، وأجب عن الأسئلة التالية:

- هل يختلف ميلا دالتي الانحدار حسب نوع المدرسة أم أنهما متساويان؟
- هل تختلف قيم نقطتي التقاطع (Intercepts) لدالتي انحدار نوع المدرسة؟
- هل تتطابق دالة انحدار درجات التلاميذ في المدارس الخاصة مع دالة انحدار درجات التلاميذ في المدارس الحكومية، أي هل لديهما تقاطع وميل واحد؟

ولبناء هذا النموذج يتم أولاً إيجاد متغير التفاعل وذلك بضرب المتغير الصوري الممثل لنوع المدرسة في ساعات الاستذكار. ويوضح الإطار رقم (٥-٣) نتائج نموذج انحدار درجات التلاميذ على ساعات الاستذكار (متغير كمي) ونوع المدرسة ومتغير التفاعل بين نوع المدرسة وساعات الاستذكار. وتشير النتائج المستعرضة بالإطار إلى أن الانحدار ككل له معنوية إحصائية (p-value =0.000). وتوضح قيم الاحتمال (p-value) المناظرة لمعاملات الانحدار أن كلاً من المتغيرين – عدد ساعات الاستذكار ونوع المدرسة – يسهمان إسهاما جوهرياً في تفسير درجات التلاميذ والتنبؤ بها (p-value =0.000) ماعدا متغير التفاعل بين هذين المتغيرين (p-value=0.25). وهذا يعني أنه لا توجد اختلافات ذات دلالة إحصائية في ميلي دالتي الانحدار حسب نوع المدرسة، وبالتالي يمكن إسقاط متغير التفاعل من النموذج. أي يمكن القول بأن أثر ساعات الاستذكار على درجات التلاميذ متماثل في المدارس الحكومية والخاصة. وتأخذ معادلة الانحدار الموفقة الصيغة التالية:

$$\hat{y}_i = 177.04 + 27.91x_i + 94.32D_i + 9.52D_ix_i$$
$$(0.000) \quad (0.000) \quad (0.000) \quad (0.253)$$

حيث إن: $\hat{y}_i$ = درجات التلميذ المقدرة، x_i = عدد ساعات الاستذكار، D_i = نوع المدرسة(١= مدرسة خاصة و ٠= مدرسة حكومية) و$D_i\,x_i$ = متغير التفاعل بين نوع المدرسة و ساعات الاستذكار؛ والأرقام بين الأقواس هي قيم الاحتمال المقابلة لمعاملات النموذج..

- **هل تختلف قيم نقطتي التقاطع (Intercepts) لدالتي انحدار نوع المدرسة؟**

ويتضح من النتائج المستعرضة بالإطار (٥-٤) أن متغير نوع المدرسة يؤثر بمستوى معنوي على درجات الطلاب. وهذا يعني أن درجات التلاميذ تختلف في المتوسط حسب نوع المدرسة، حيث يصل الفرق في المتوسط إلى ٩٤,٣ درجة مما يعني أن نقطتي التقاطع تختلف بنفس هذه القيمة.

هل تتطابق دالة انحدار درجات التلاميذ للمدارس الخاصة مع دالة انحدار المدارس الحكومية، أي هل لديهما تقاطع وميل واحد؟

كما أوضحنا أن نقطتي التقاطع لدالتي الانحدار تختلفان بمستوى معنوي مما يعني أن دالتي الانحدار لا تتطابقان. ولكن بهدف الشرح والتوضيح نقوم باختبار تطابق دالتي الانحدار كما يلي:

لإجراء هذا الاختبار يستخدم اختبار F الجزئي المتعدد وذلك ببناء نموذج انحدار كامل يضم متغيرات ساعات الاستذكار (X_i)، نوع المدرسة الكمي (D_i)، متغير التفاعل بينهما (D_iX_i) ونموذج انحدار مخفض يضم متغير ساعات الاستذكار (X_i). والفرض المراد اختباره في هذه الحالة هو:

فرض العدم: ($H_0: \beta_2=\beta_3=0$) في مقابل الفرض البديل: ليست كلتا قيمتي المعلمتين β_2 و β_3 مساوية للصفر

ومن نتائج النموذجين تم الحصول على القيم التالية:

مجموع مربعات انحدار النموذج الكامل = ٢٧١٩٧٩,٣٨٤٢

مجموع مربعات انحدار النموذج المخفض = ١٢٠٥٣٥,٩

مجموع مربعات البواقي للنموذج الكامل = ٤٨١٤٠,٨٩٤٨٦

والآن يتم حساب قيمة إحصاء F_0 كما يلي:

$$F_o = \frac{(271979.3842 - 120535.9)/2}{48140.89486/39} = 61.3434$$

وبما أن قيمة F_0 المحسوبة أكبر بكثير من قيمة توزيع F بدرجتي حرية ٢ و ٣٩ ($F(0.05;2;39)=3.24$) ، فإننا نرفض فرض العدم ونحكم بعدم تطابق خطي الانحدار.

إطار رقم (٥-٣): نتائج نموذج انحدار درجات التلاميذ على ساعات الاستذكار (متغير كمي) ونوع المدرسة ومتغير التفاعل بين نوع المدرسة وساعات الاستذكار.

Regression Statistics	
Multiple R	0.921746293
R Square	0.849616229
Adjusted R Square	0.838048246
Standard Error	35.13377178
Observations	43

ANOVA

	Df	SS	MS	F	Significance F
Regression	3	271979.3842	90659.79	73.4455	0.0000
Residual	39	48140.89486	1234.382		
Total	42	320120.2791			

	Coefficients	Standard Error	t Stat	P-value
Intercept	177.040146	16.30329925	10.85916	0.0000
نوع المدرسة	94.31699687	24.53875123	3.843594	0.0004
عدد ساعات الاستذكار	27.9136253	5.876960681	4.74967	0.0000
متغير التفاعل	9.514946124	8.197211305	1.160754	0.2528

٥-٣-٥ نموذج انحدار يشتمل على متغير كمي ومتغيرين نوعيين:

من الممكن بناء نموذج انحدار يضم متغيرين نوعيين أو أكثر. فعلى سبيل المثال يمكن أن يضم نموذج الانحدار متغيرين نوعيين لكل منهما صفتين ومتغير كمي واحد. وفي هذه الحالة يأخذ نموذج الانحدار بدون إدخال متغيرات تفاعل الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_{1i} + \beta_3 D_{2i} X_i + \varepsilon_i$$

حيث إن:

Y_i = المتغير التابع.

X_i = متغير كمي.

D_{1i} = متغير صوري يأخذ القيمة "١" إذا كانت المشاهدة تنتمي للصفة الأولى والقيمة "صفر" إذا كانت المشاهدة تنتمي إلى الصفة الثانية للمتغير النوعي الأول.

D_{2i} = متغير صوري يأخذ القيمة "١" إذا كانت المشاهدة تنتمي للصفة الأولى والقيمة "صفر" إذا كانت المشاهدة تنتمي إلى الصفة الثانية للمتغير النوعي الثاني.

ε_i = حد الخطأ العشوائي.

ويوضح هذا النموذج أن دوال الانحدار المناظرة لصفات المتغيرات النوعية لها ميل ثابت ونقاط تقاطع مختلفة. وفي حالة إدخال متغيرات تفاعل يأخذ النموذج الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 D_{1i} X_i + \beta_5 D_{2i} X_i + \beta_6 D_{1i} D_{2i} + \varepsilon_i$$

ويلاحظ أن هذا النموذج يضم ثلاثة متغيرات تفاعل؛ أي ما يعادل توفيقات اختيار متغيرين من بين ثلاثة متغيرات. ويتم إيجاد القيم المتوقعة للمتغير التابع كما يلي:

المتغير النوعي الأول(D_1)	المتغير النوعي الثاني (D_2)	القيم المتوقعة للمتغير التابع
الصفة الأولى (1)	الصفة الأولى (1)	$E(Y_i) = (\beta_0+\beta_2+\beta_3+\beta_6) + (\beta_1+\beta_4+\beta_5)X_i$
الصفة الثانية (0)	الصفة الأولى (1)	$E(Y_i) = (\beta_0+\beta_3) + (\beta_1+\beta_5)X_i$
الصفة الأولى (1)	الصفة الثانية (0)	$E(Y_i) = (\beta_0+\beta_2) + (\beta_1+\beta_4)X_i$
الصفة الثانية (0)	الصفة الثانية (0)	$E(Y_i) = \beta_0+\beta_1 X_i$

وفي مثل هذا النموذج ينصب الاهتمام على إجابة الأسئلة التالية:

- هل هناك تأثير تفاعل بين المتغير النوعي الأول (D_1) والمتغير الكمي (X_i) له دلالة إحصائية؟
- هل هناك تأثير تفاعل بين المتغير النوعي الثاني (D_2) والمتغير الكمي (X_i) له دلالة إحصائية؟
- هل هناك تأثير تفاعل بين المتغيرين النوعين (D_1) و(D_2) له دلالة إحصائية؟

ويستخدم إحصاء F الجزئي أو اختبار t للإجابة عن هذه الأسئلة. ويتعين على الباحث اختبار معنوية متغيرات التفاعل أولاً، وفي حالة عدم معنويتها يتم اختبار وتفسير المتغيرات الأساسية المكونة لمتغيرات التفاعل.

كما يمكن اختبار توازي دوال الانحدار الأربعة؛ وفي هذه الحالة فإن الفرض المراد اختباره هو:

فرض العدم $(H_0: \beta_4 = \beta_5 = 0)$ **في مقابل الفرض** البديل : $(H_1: \beta_4 \neq 0 \text{ or } \beta_5 \neq \beta_4 \text{ or } \beta_5 \neq 0)$

ولإجراء هذا الاختبار نقوم أولاً ببناء نموذج انحدار كامل يضم كل المتغيرات الستة ونموذج انحدار مخفض يضم المتغيرات الأربعة (X_i, D_1, D_2, D_1D_2) ومن ثم يتم حساب إحصاء F_0 حيث

$$F_0 = \frac{[ESS(X_i, D_{1i}, D_{2i}, D_{1i}X_i, D_2X_i, D_{1i}D_{2i}) - ESS(X_i, D_{1i}, D_{2i}, D_{1i}D_{2i})]/2}{RSS(X_i, D_{1i}, D_{2i}, D_{1i}X_i, D_2X_i, D_1D_2)/(n-7)}$$

فإذا كانت قيمة F_0 أكبر من قيمة توزيع F بدرجتي حرية 2 و(n-7) وعند مستوى معنوية معين (α) نرفض فرض العدم ونحكم بأن دوال الانحدار متوازية وبنفس الطريقة يمكننا إجراء اختبار تطابق دوال الانحدار وفي هذه الحالة فإن الفرض المراد اختباره هو:

فرض العدم: $(H_0: \beta_2=\beta_3=\beta_4=\beta_5=\beta_6=0)$ **في مقابل الفرض البديل ليس كل قيم هذه المعلمات مساوية للصفر**

ولإجراء هذا الاختبار يتم بناء نموذج انحدار كامل يضم كل المتغيرات الستة ونموذج انحدار مخفض يضم المتغير الكمي فقط ويتم حساب إحصاء F_0 حيث

$$F_0 = \frac{[ESS(X_i, D_{1i}, D_{2i}, D_{1i}X_i, D_{2i}X_i, D_{1i}D_{2i}) - ESS(X_i)]/5}{ESS(X_i, D_{1i}, D_{2i}, D_{1i}X_i, D_{2i}X_i, D_{1i}D_{2i})/(n-7)}$$

فإذا كانت قيمة F_0 أكبر من قيمة توزيع F بدرجتي حرية 5 و (n-7) وعند مستوى معنوية معين (α) نرفض فرض العدم ونحكم بعدم تطابق دوال الانحدار، وأما إذا كانت قيمة F_0 أقل من قيمة توزيع F نحكم بأن دوال الانحدار متطابقة، أي يمكننا استخدام نموذج الانحدار التالي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

مثال (٥-٥):

من بيانات المثال السابق أجرِ نموذج انحدار درجات التلميذ على ساعات الاستذكار ونوع المدرسة ومستوى تعليم رب الأسرة ومتغيرات التفاعل نوع المدرسة ومستوى تعليم رب الأسرة، نوع المدرسة وساعات الاستذكار، ومستوى تعليم رب الأسرة وساعات الاستذكار وأجب عن الأسئلة التالية:

- هل هناك تأثير تفاعل بين متغير نوع المدرسة (D_1) ومتغيري تعليم رب الأسرة (D_2) و (D_3) على درجات التلميذ له دلالة إحصائية؟.
- هل هناك تأثير تفاعل بين متغير نوع المدرسة (D_1) وساعات الاستذكار (X_i) على درجات التلميذ له دلالة إحصائية؟.
- هل هناك تأثير تفاعل بين متغير تعليم رب الأسرة ((D_2) و(D_3)) وساعات الاستذكار (X_I) على درجات التلميذ له دلالة إحصائية؟.

الحل:

يعد هذا المثال أكثر تعقيداً من المثال النظري الذي تم شرحه في الجزء (٥-٣-٥) إذ يضم نموذج الانحدار متغيرين نوعيين أحدهما يتضمن فئتين والآخر يحتوي على ثلاث فئات. ولبناء النموذج يتم أولاً تعريف متغيرات التفاعل التالية:

نوع المدرسة (D_1) ومستوى تعليم رب الأسرة (D_2) (١=غير متعلم ، ٠= بخلاف ذلك)	$D_2 \times D_1$
نوع المدرسة (D_1) ومستوى تعليم رب الأسرة (D_3) (١=تعليم متوسط ، ٠= بخلاف ذلك)	$D_3 \times D_1$
نوع المدرسة (D_1) ومتوسط عدد ساعات الاستذكار(X)	$X \times D_1$
مستوى تعليم رب الأسرة (D_2) (١=غير متعلم ، ٠= بخلاف ذلك) و ساعات الاستذكار (X)	$X \times D_2$
مستوى تعليم رب الأسرة (D_3) (١=تعليم متوسط ، ٠= بخلاف ذلك) و ساعات الاستذكار(X)	$X \times D_3$

بإدخال متغيرات التفاعل أصبح النموذج يضم (٩) متغيرات؛ (٤) متغيرات أساسية و (٥) متغيرات تفاعل.

ويأخذ النموذج الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_3 \beta_5 D_1 D_2 + \beta_6 D_1 D_3 + \beta_7 D_1 X + \beta_8 D_2 X + \beta_9 D_3 X + \varepsilon_i$$

ويوضح الجدول التالي القيم المتوقعة لدرجات التلميذ حسب نوع المدرسة ومستوى تعليم رب الأسرة:

قيم المتغيرات			القيم المتوقعة لدرجات التلاميذ المتوقعة	ملاحظات
D_1	D_2	D_3		
1	0	0	$Y = (\beta_0+\beta_2) +(\beta_1+\beta_7)X$	مدرسة خاصة ومستوى تعليم والده "ثانوي فما فوق"
1	1	0	$Y=(\beta_0+\beta_2+\beta_3+\beta_5) +(\beta_1+\beta_7+\beta_8)X$	مدرسة خاصة ومستوى تعليم والده "غير متعلم"
1	0	1	$Y=(\beta_0+\beta_2+\beta_4+\beta_6) +(\beta_1+\beta_7+\beta_9)X$	مدرسة خاصة ومستوى تعليم والده "تعليم متوسط"
0	0	0	$Y = \beta_0+\beta_1X$	مدرسة حكومية ومستوى تعليم والده "ثانوي فما فوق"
0	1	0	$Y = (\beta_0+\beta_3)+(\beta_1 +\beta_8)X$	مدرسة حكومية ومستوى تعليم والده "غير متعلم"
0	0	1	$Y = (\beta_0+\beta_4) +(\beta_1+\beta_9)X$	مدرسة حكومية ومستوى تعليم والده "تعليم متوسط

وتشير النتائج المستعرضة بالإطار(٥-٥) إلى معنوية الانحدار ككل (p-value = 0.000) . وتفسر المتغيرات المضمنة في النموذج مجتمعة ٩٧,٦% من التغير في درجات التلاميذ. وكما سبق ذكره أنه يجب على الباحث اختبار معنوية متغيرات التفاعل أولاً وفي حالة عدم معنويتها يتم اختبار وتفسير المتغيرات الأساسية المكونة لمتغيرات التفاعل. أي هل دوال الانحدار الموضحة أعلاه لها ميل واحد؟ أم هناك اختلاف في كل أو بعض ميول هذه الدوال؟ الفرض المراد اختباره في هذه الحالة هو:

فرض العدم: $(H_0 =\beta_7=\beta_8=\beta_9 = 0)$ **في مقابل** الفرض البديل ليس كل قيم هذه المعلمات مساوية للصفر.

ولإجراء هذا الاختبار يستخدم اختبار F الجزئي المتعدد وذلك ببناء نموذج انحدار كامل يضم المتغيرات التسعة ونموذج مخفض يضم جميع المتغيرات ما عدا متغيرات التفاعل: D_1X, D_2X, D_3X. ومن نتائج النموذجين تم الحصول على القيم التالية:

مجموع مربعات انحدار النموذج الكامل = ٣١٢٥١٢,٠٣

مجموع مربعات انحدار النموذج المخفض = ٣١٠٣٠٨,٩٨

مجموع مربعات البواقي للنموذج الكامل = ٧٦٠٨,٢٥

والآن يتم حساب قيمة إحصاء F_0 كما يلي:

$$F_0 = \frac{(312512.03 - 310308.98)/3}{7608.25/33} = 9.56$$

وبما أن قيمة F_0 المحسوبة أكبر بكثير من قيمة توزيع F بدرجتي حرية 3 و33 $(F(0.05;3;33)=2.89)$ ، فإننا نرفض فرض العدم ونحكم بوجود تأثير تفاعل بين المتغيرين النوعيين والمتغير الكمي. وبعبارة أخرى أن دوال الانحدار ليس لديها ميل واحد أو غير متوازية. وتوضح النتائج المستعرضة بالإطار أن متغيرات التفاعل $(D_1D_2,\ D_1D_3,\ D_2X,\ D_3X)$ ذات دلالة إحصائية $(p\text{-value} < 0.01)$. ويستشف من هذه النتائج أن هناك تأثيراً مشتركاً بين نوع المدرسة ومستوى تعليم رب الأسرة على درجات التلميذ وكذلك يوجد تأثير مشترك بين عدد ساعات الاستذكار ومستوى تعليم رب الأسرة. وهذا يشير إلى أن هناك اتجاهاً لدى بعض الآباء المتعلمين لإلحاق أبنائهم بمدارس خاصة وأن متوسط ساعات الاستذكار التلميذ يختلف باختلاف مستوى تعليم رب الأسرة. كما تشير النتائج إلى عدم وجود تأثير تفاعل بين نوع المدرسة وساعات الاستذكار له دلالة إحصائية $(p\text{-value}=0.235)$.

إطار رقم (٥-٥): نتائج نموذج انحدار درجات التلاميذ على متغيرات نوع المدرسة ومستوى تعليم رب الأسرة ومتوسط ساعات الاستذكار ومتغيرات التفاعل

Regression Statistics	
Multiple R	0.9880
R Square	0.9762
Adjusted R Square	0.9698
Standard Error	15.1840
Observations	43

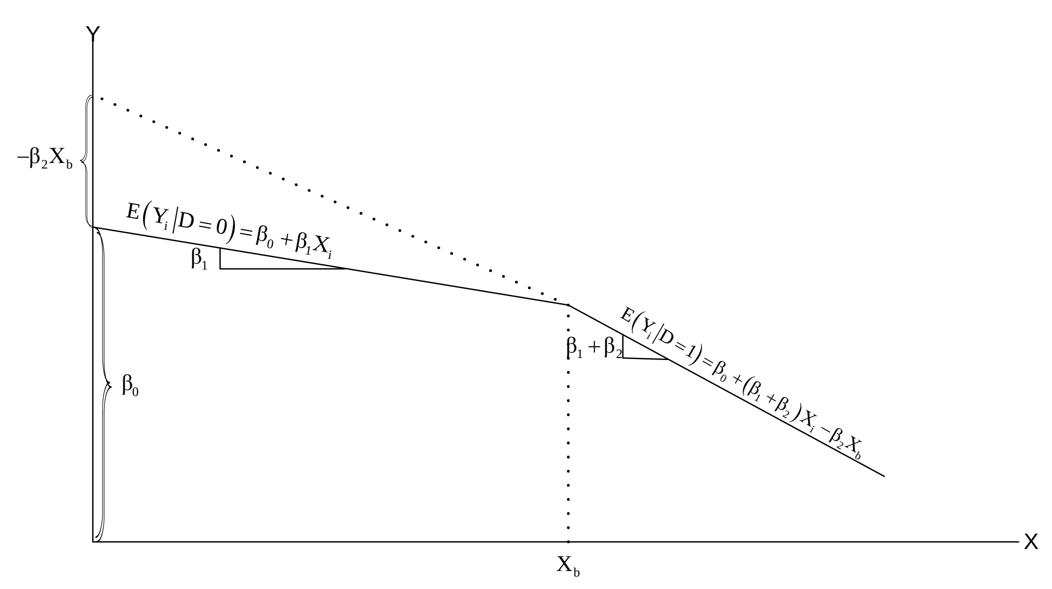
ANOVA

	df	SS	MS	F	Significance F
Regression	9	312512.0313	34723.5590	150.6099	0.0000
Residual	33	7608.2478	230.5530		
Total	42	320120.2791			

	Coefficients	Standard Error	t Stat	P-value
Intercept	164.3508	12.8951	12.7453	0.0000
D_1	126.5242	13.8991	9.1031	0.0000
X	41.9895	5.6741	7.4002	0.0000
D_2	8.9351	14.9394	0.5981	0.5539
D_3	19.6581	14.3465	1.3702	0.1799
D_1D_2	-57.0483	14.3985	-3.9621	0.0004
D_1D_3	40.9073	15.4133	2.6540	0.0121
D_1X	-5.4235	4.4797	-1.2107	0.2346
D_2X	-17.5486	5.9915	-2.9289	0.0061
D_3X	-13.5443	5.2063	-2.6015	0.0138

٤-٥ نموذج الانحدار الخطي القطعي* (Piecewise linear regression model)

يمكن نمذجة العلاقة غير الخطية بين المتغير التابع والمتغير المستقل باستخدام نموذج خطي يحتوي على ميول (slopes) مختلفة لنطاقات (أمدية) محددة لقيم المتغير المستقل. ويسمى النموذج الخطي في هذه الحالة بنموذج الانحدار الخطي القطعي. فمثلاً قد تكون العلاقة بين المتغير التابع ومدى محدد من قيم المتغير المستقل علاقة طردية وفي مدى آخر لقيم المتغير المستقل تكون العلاقة عكسية. ويستخدم النموذج القطعي عندما يكون خط الانحدار مقسماً إلى عدد من قطع الخطوط مفصولة بواسطة عقد (knots) يتغير عندها اتجاه الانحدار (شكل ٥-١٠).



شكل (٥-١٠): شكل نموذج الانحدار الخطي القطعي

وتستخدم المتغيرات الصورية لتمثل عدد القطع التي تتغير عندها العلاقة بين المتغير التابع والمستقل. وفي حالة وجود عقدة تغير واحدة (X_b) - بتحديد العقدة التي يتغير عندها ميل الانحدار - يأخذ نموذج الانحدار بافتراض أن دالة الانحدار مستمرة عن X_b الصيغة التالية (Eye and Schuster, 1998):

$$Y_i=\beta_0+\beta_1 X_i+\beta_2(X_i-X_b)D+\varepsilon_i \qquad (5\text{-}9)$$

حيث إن:

Y_i المتغير التابع.

X_i المتغير المستقل.

X_b النقطة (قيمة المتغير المستقل) التي يتغير عندها ميل الانحدار.

* للمزيد حول نماذج الانحدار القطعي والنماذج الشرائحية (Spline models) يرجى الرجوع إلى (Montgomery et al 2001, Draper and Smith, 1998)

D متغير صوري يأخذ القيمة "١" إذا كانت قيمة المتغير المستقل أكبر من X_b والقيمة "صفر" إذا كانت قيمـة المتغـير المستقل أصغر من أو تساوي X_b.

ε_i حد الخطأ العشوائي

من المعادلة (5.9) يمكن حساب القيمة المتوقعة للمتغير التابع إذا كانت قيمة المتغير المستقل أقل من أو تساوي X_b ، أي عندما تكون (D=0):

$$E\left(Y_i \mid D=0\right)=\beta_0+\beta_1 X_i$$

وحساب القيمة المتوقعة للمتغير التابع إذا كانت قيمة المتغير المستقل أكبر من X_b ، أي عندما تكون (D=1):

$$E\left(Y_i \mid D=1\right)=\left(\beta_0-\beta_2 X_b\right)+\left(\beta_1+\beta_2\right) X_i$$

ومن ثم فإن β_1 و $\left(\beta_1+\beta_2\right)$ هما ميلا خطي الانحدار و β_0 و $\left(\beta_0-\beta_2 X_b\right)$ هما ثابتا المتغير التابع (الأجزاء المقطوعة من المتغير التابع).

مثال:

يوضح الجدول (٥-٣) بيانات افتراضية عن العلاقة بين إنتاجية محصول القمح (كجم/هكتار) وكميات سماد النيتروجـين المستخدمة (كجم/هكتار) لتجربة زراعية عن أثر إضافة السماد في إنتاجية المحصول. ويتضح مـن الشـكل رقـم (٥-١١) أن العلاقة بين الإنتاجية وكمية السماد المستخدمة طردية إلى أن تصل كمية السماد (٨٠ كجم/هكتار) ومـن ثـم تبـدأ العلاقـة عكسية بين الإنتاجية وكمية السماد. لذا فإن النموذج الملائم في هذه الحالة هو النموذج القطعي عند النقطـة أو الجرعـة (٨٠ كجم / هكتار). ويأخذ النموذج المراد توفيقه الصيغة التالية:

$$Y_i=\beta_0+\beta_1 X_i+\beta_2\left(X_i-80\right) D+\varepsilon_i$$

حيث إن:

Y_i إنتاجية القمح (كجم/هكتار).

X_i كمية السماد (كجم/هكتار).

٨٠ جرعة السماد التي يتغير عندها ميل الانحدار.

D متغير صوري يأخذ القيمة "١" إذا كانت جرعة السماد أكبر من ٨٠ كجم/هكتار والقيمة "صفر" إذا كانت الجرعـة أقل من أو تساوي ٨٠ كجم/هكتار.

ε_i حد الخطأ العشوائي

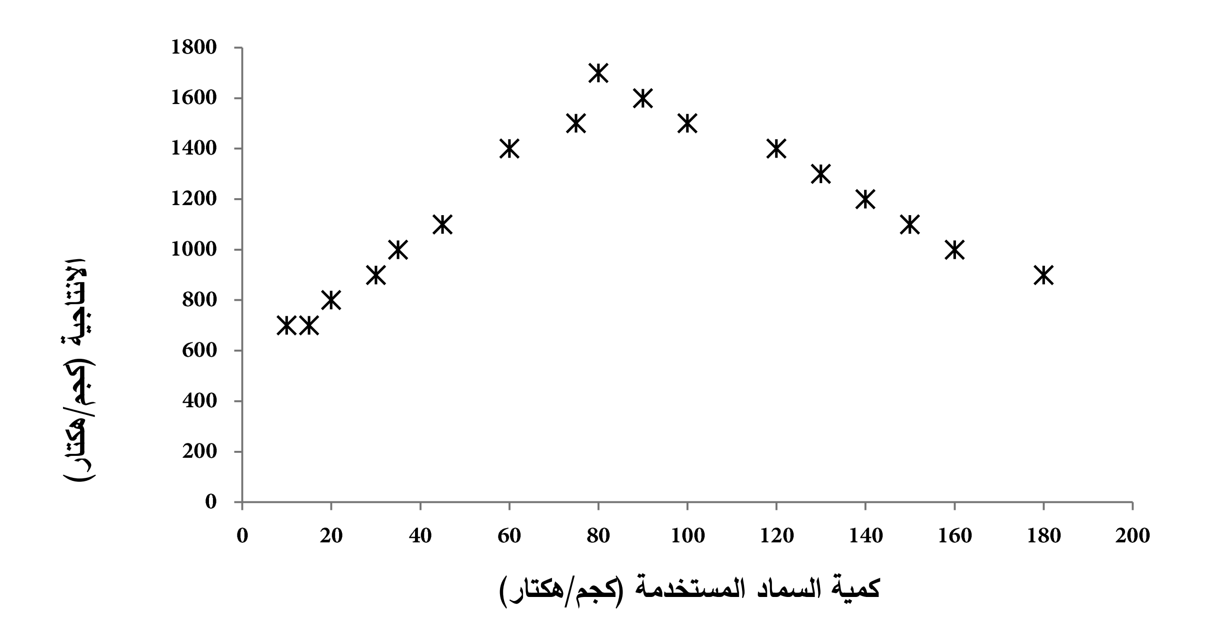
ولإجراء الانحدار القطعي تمَ أولاً حساب المتغير $(X_i-80)\times D$ ، ومن ثم حصلنا على النتـائج الموضـحة بالاطـار رقـم (٥-٦). ومن النتائج يتضح أن زيادة واحد كيلوجرام سماد للهكتـار تسـهم في زيـادة الإنتاجيـة بمقـدار (١٢,٥) كجـم قمـح

للهكتار إذا استمرت الإضافة إلى (٨٠) كجم/للهكتار. وإذا استمرت الإضافة بأكثر من (٨٠) كيلـوجرام للهكتـار، فـإن إضـافة كل كيلوجرام سماد ستسهم في انخفاض إنتاجية محصول القمح بمقدر (٩,٢٨) كجم/الهكتار (١٢,٤٦-٢١,٧٤= -٩,٢٨).

جدول رقم (٥-٣): إنتاجية محصول القمح وكميات السماد النيتروجيني المستخدمة

المشاهدة	الانتاجية (كجم/هكتار) Y_i	جرعة السماد (كجم/هكتار) X_i	$(X_i - 80) \times D$
1	700	10	0
2	1000	35	0
3	1100	45	0
4	1400	120	30
5	900	180	90
6	1300	130	40
7	1500	75	0
8	1400	60	0
9	1500	100	10
10	1700	80	0
11	1100	150	60
12	1200	140	50
13	1000	160	70
14	900	30	0
15	1600	90	0
16	800	20	0
17	700	15	0

مصدر: بيانات افتراضية



شكل (٥-١١): شكل إنتاجية القمح مع كمية السماد المستخدمة

إطار رقم (٥-٦): نتائج انحدار إنتاجية محصول القمح على كمية السماد المستخدمة

Regression Statistics	
Multiple R	0.983471
R Square	0.967216
Adjusted R Square	0.962532
Standard Error	61.18857
Observations	17

ANOVA

	df	SS	MS	F	Significance F
Regression	2	1546406.95	773203.47	206.52	0.0000
Residual	14	52416.58	3744.04		
Total	16	1598823.53			

	Coefficients	Standard Error	t Stat	P-value
Intercept	556.80	36.58	15.22	0.0000
جرعة السماد	12.46	0.62	20.11	0.0000
$(X_i - 80) \times D$	-21.74	1.13	-19.28	0.0000

٥-٥ ملاحظات:

- في حالة إدخال متغيرات تفاعل في نموذج الانحدار يجب أن يضم النموذج المتغيرات الأساسية التي استخدمت لإيجاد هذه المتغيرات اتساقاً مع مبدأ التهميش (Principle of Marginality) . فمثلاً النموذج التالي لا يوافق مبدأ التهميش:

$$Y_i=\beta_0+\beta_1X_i+\beta_3D_iX_i+\varepsilon_i$$

لأن النموذج لا يضم المتغير الصوري (D) في حين تم إدخاله في متغير التفاعل ($D_i X_i$) .

- لا يتم عادةً اختبار وتفسير المتغيرات الأساسية في حالة وجود متغير تفاعل متكون من هذه المتغيرات له دلالة إحصائية. وهذا يعني أنه يتم أولاً اختبار متغيرات التفاعل وفي حالة عدم معنويتها يتم اختبار وتفسير المتغيرات الأساسية المكونة لها.

- في حالة بناء نموذج انحدار معياري (Standardized regression model)، ينصح بأن لا يتم تحويل المتغيرات الصورية إلى متغيرات معيارية وذلك لصعوبة تفسيرها لأن المتغير لا يتغير بانحراف معياري واحد.

- يتعين على الباحث أن يستخدم طرق اختيار المتغيرات في نموذج الانحدار – موضوع الفصل السادس- بحذر في حالة اشتمال النموذج على متغيرات تفاعل، ذلك لأن هذه الطرق تختار المتغيرات التي تدخل النموذج بصورة ذاتية حسب مساهمتها في تفسير تباين المتغير التابع، والتي ربما تقود إلى اختيار متغيرات تفاعل وإسقاط المتغيرات المكونة لها.

تمارين

١. البيانات التالية تم الحصول عليها من عينة عشوائية قوامها ٢٥ موظفاً يعملون في شركة ما.

رقم المشاهدة	الراتب السنوي (ألف دولار)	سنوات الخدمة	النوع
١	٣٥	١٥	رجل
٢	٢٧	١٧	امرأة
٣	٤٥	٢٥	رجل
٤	٢٢	١٢	امرأة
٥	٢٥	٢	رجل
٦	٣٠	١٠	رجل
٧	٣٧	١٧	رجل
٨	٢٥	١٧	امرأة
٩	١٧	١	امرأة
١٠	٢٨	٤	رجل
١١	٤٣	٢٥	رجل
١٢	٢٥	١٥	امرأة
١٣	٢٢	١	رجل
١٤	٢٨	٦	رجل
١٥	٢٩	٢٠	امرأة
١٦	١٩	٣	امرأة
١٧	٢٩	٢١	امرأة
١٨	٣٨	١٩	رجل
١٩	١٩	٥	امرأة
٢٠	٢٢	١	رجل
٢١	٣٩	٢٠	رجل
٢٢	٤٠	٢٢	رجل
٢٣	٢١	١٠	امرأة
٢٤	٢٨	٧	رجل
٢٥	٣٠	٨	رجل

المطلوب:

- إجراء انحدار الراتب على سنوات الخبرة والنوع؟
- هل معدل الزيادة في الراتب حسب الخبرة متساو بين الرجال والنساء؟
- هل بداية الرواتب متساوية بين الرجال والنساء؟
- قدر متوسط راتب النساء اللائي أكملن ١٦ سنة في الخدمة.

٢. البيانات المستعرضة بالجدول التالي بيانات افتراضية عن بعض العوامل المؤثرة على الرضا الوظيفي.

رقم المشاهدة	درجات الرضا من (١٠٠) درجة	**القطاع**	الراتب (ألف ريال)	النوع	مستوى التعليم
1	66	عام	6.6	رجل	متوسطة
2	45	خاص	4	رجل	متوسطة
3	56	خاص	5	امرأة	متوسطة
4	51	خاص	5	رجل	ثانوى
5	62	خاص	5.6	امرأة	جامعى
6	80	عام	7.1	رجل	جامعى
7	67	خاص	6.4	امرأة	جامعى
8	81	عام	7	امرأة	جامعى
9	79	عام	7.9	رجل	جامعى
10	63	خاص	6.2	رجل	جامعى
11	79	عام	6.9	رجل	ثانوى
12	78	عام	6.5	امرأة	جامعى
13	77	عام	6	امرأة	ثانوى
14	66	خاص	6	امرأة	ثانوى
15	71	عام	6	امرأة	جامعى
16	61	عام	5.1	رجل	متوسطة
17	47	خاص	4	رجل	متوسطة
18	75	عام	6	رجل	ثانوى
19	41	خاص	4	رجل	ثانوى
20	39	خاص	3.9	امرأة	ثانوى
21	63	عام	6	امرأة	ثانوى
22	61	عام	6.1	رجل	متوسطة
23	73	عام	7.3	رجل	متوسطة
24	49	خاص	4.2	امرأة	ثانوى
25	57	عام	5.7	امرأة	متوسطة
26	59	خاص	5.2	امرأة	جامعى
27	73	عام	6.3	امرأة	جامعى
28	65	عام	5.5	امرأة	ثانوى
29	74	عام	7.4	امرأة	جامعى
30	71	عام	7	رجل	ثانوى
31	62	عام	5.2	رجل	ثانوى
32	67	خاص	6.3	رجل	متوسطة

المطلوب:

- تحويل المتغيرات النوعية إلى متغيرات صورية.
- إجراء انحدار الرضا الوظيفي على القطاع، والراتب ، والنوع، ومستوى التعليم؟
- إجراء انحدار الرضا الوظيفي على القطاع، الراتب، النوع ومستوى التعليم ومتغيرات التفاعل بين هذه المتغيرات؟
- فسر النتائج التي تحصل عليها؟

الفصل السادس
اختيار المتغيرات المستقلة في نموذج الانحدار الخطي

٦-١ مقدمة:

من الصعوبات التي تواجه الباحثين في بناء نموذج الانحدار الخطي المتعدد اختيار المتغيرات المستقلة التي تدخل في النموذج لتعطي أعلى درجة من الدقة في التنبؤ بالمتغير التابع. وفي بعض الحقول كالعلوم الطبيعية والزراعية تساعد النظرية في اختيار المتغيرات المستقلة التي تدخل في نموذج الانحدار ويتم أحياناً في مثل هذه الحقول تصميم تجارب تُحدد فيها المتغيرات المستقلة، إلا أنه في بعض الحقول الأخرى كالعلوم الاجتماعية والإنسانية نادراً ما نجد نماذج نظرية تُحدد بشكل قاطع المتغيرات المستقلة التي تسهم في تفسير تباين المتغير التابع محل الدراسة. وعادةً ما يكون لدى الباحث عدد كبير من المتغيرات المرشحة للدخول في نموذج الانحدار. إلا أن بعض هذه المتغيرات ربما لا تسهم بمستوى دال إحصائياً في عملية التنبؤ. ولذلك من المفيد عملياً التركيز على عدد قليل من المتغيرات المهمة شريطة أن يعطي هذا العدد معادلة انحدار لا تقل كفاءةً عن المعادلة التي تستخدم جميع المتغيرات المتاحة. ويقترح جاتفيلد (Chatfield, 1995, p257) في هذا الصدد ألا يزيد عدد المتغيرات المستقلة عن رُبع عدد المشاهدات ويفضل ألا يزيد عددها عن أربعة أو خمسة متغيرات. ومن أهداف تقليل عدد معالم نموذج الانحدار الخطي المتعدد نذكر ما يلي (انظر أيضاً Neter 1990 p 436):

- التخلص من بعض المتغيرات التي لا تسهم في تفسير تباين المتغير التابع.
- وجود عدد كبير من المتغيرات المستقلة لا يناظرها عدد مناسب من المشاهدات.
- تجنب مشكلة الارتباط الخطي المتعدد بين المتغيرات المستقلة (Multicollinearity) والتي غالباً ما تبرز عندما يتم قياس متغيرات عديدة من أفراد العينة.
- سهولة فهم واستخدام النموذج ذي المعالم القليلة.

يعالج هذا الفصل موضوع "اختيار أفضل نموذج انحدار" عندما تكون لدى الباحث مجموعة كبيرة من المتغيرات المستقلة بهدف الوصول إلى نموذج انحدار يضم عدداً قليلاً من هذه المتغيرات ويعطي أعلى درجة من الدقة في التنبؤ بالمتغير التابع.

٦-٢ معايير اختيار أفضل نموذج:

يتطلب اختيار النموذج الأفضل تحديد معيار أو معايير للمفاضلة بين النماذج المرشحة. لذا يعد تحديد المعيار الخطوة الأولى في اختيار أفضل نموذج. ومعيار الاختيار هو مقياس يتم حسابه لكل نموذج مرشح ويستخدم للمفاضلة بين هذه النماذج. وباستخدام معيار محدد يمكننا ترتيب النماذج المرشحة من الأفضل إلى الأسوأ. وتجدر الإشارة هنا إلى أنه تتم المقارنة بين النماذج فقط عندما يكون المتغير التابع لكل نموذج واحد.

ومن المعايير الواسعة الاستخدام للمفاضلة بين النماذج مقياس معامل التحديد R^2، إحصاء F ، الخطأ المعياري للتقدير s أو التباين s^2 وإحصاء ملاوس (C_p) وإحصاء مجموع مربعات التنبؤ (PRESS) معايير المعلومات AIC و SBCوBIC ومعيار التنبؤ لأمميايا Amemiaya Prediction Criteria (PC). ويرجع استخدام أكثر من معيار للمفاضلة إلا أنه لا يوجد معيار واحد باستخدامه نحصل دائماً على النموذج الأفضل. وتستخدم هذه المقاييس للمقارنة بين معادلتي

انحدار، أحدهما يعرف بالنموذج الكامل (Full Model) ويضم عدد (k) متغير مستقل والآخر يعرف بالنموذج المخفض (Reduced Model) ويحتوى على عدد (p) من المتغيرات المستقلة، حيث p أقل من k، أي أن عدد متغيرات النموذج الكامل يزيد عن عدد متغيرات النموذج المخفض بعدد (k-p) متغير. ويأخذ النموذج الكامل الصيغة التالية:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \beta_{p+1} x_{p+1} + \ldots + \beta_k x_k + \varepsilon \qquad (6\text{-}1)$$

ويأخذ النموذج المخفض الصيغة التالية:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \qquad (6\text{-}2)$$

٦-٢-١ المعيار الأول - معامل التحديد R^2:

يستخدم معامل التحديد R^2 كمعيار للمفاضلة بين نموذجي انحدار أو أكثر عندما يكون المتغير التابع واحداً. فالنموذج الذي له أعلى قيمة R^2 هو الذي يفسر أكبر تباين في المتغير التابع. إلا أنه من أهم عيوب معامل التحديد - كما أشرنا إلى ذلك في الفصل الثالث - أن قيمته تزيد بإضافة أي متغير مستقل لنموذج الانحدار حتى لو كان هذا المتغير لا يسهم بمستوى معنوي في تفسير تباين المتغير التابع. ولذلك نجد أن قيمة معامل التحديد R^2 للنموذج الكامل دائماً أكبر من قيمة R^2 للنموذج المخفض لاحتواء الأول على متغيرات مستقلة أكثر من الثاني. ولتجنب هذا العيب يفضل استخدام معامل التحديد المعدل $\bar{R}^2$ كمعيار للمفاضلة بين النماذج المرشحة.

٦-٢-٢ المعيار الثاني - إحصاء F الجزئي:

يستخدم اختبارF الجزئي للتعرف على دلالة ما يحدث من نقص في دقة التنبؤ عندما يستبعد واحد أو أكثر من المتغيرات المستقلة من نموذج الانحدار. وبصفة عامة إذا تم بناء نموذج انحدار يضم (k) متغير مستقل وبناء نموذج آخر يضم (p) متغيراً مستقلاً، حيث (p) أقل من (k)، أي بعد استبعاد (k-p) وأردنا اختبار ما إذا كانت دقة التنبؤ قد انخفضت انخفاضاً ذا دلالة إحصائية نتيجة للاستبعاد فإننا نستخدم الاختبار التالي:

$$F_p = \frac{[RSS(p) - RSS(k)]/(k-p)}{RSS(k)/(n\text{-}k\text{-}1)} \sim F_{(k\text{-}p),(n\text{-}k\text{-}1)} \qquad (6.3)$$

حيث إن:

k = عدد المتغيرات المستقلة في النموذج الكامل.

p = عدد المتغيرات المستقلة في النموذج المخفض.

n = عدد المشاهدات.

RSS(p) = مجموع مربعات البواقي للنموذج المخفض.

RSS(k) = مجموع مربعات البواقي للنموذج الكامل.

وتتبع الإحصاءة F_p توزيع F بدرجتي حرية (k-p) و(n-k-1). ويستخدم هذا المعيار لاختبار ما إذا كان الفرق بين مجموع مربعات البواقي للنموذج المخفض ومجموع مربعات بواقي النموذج الكامل يختلف عن الصفر أم لا؟. فإذا كان

الفرق غير معنوي فإن بالإمكان استخدام النموذج المخفض للحصول على نفس القدرة التنبئية تقريباً التي يمكن الحصول عليها إذا ما استخدمنا النموذج الكامل.

٦-٢-٣ المعيار الثالث - مقدر الانحراف المعياري (s):

يستخدم الانحراف المعياري (s) أو التباين (s^2) كمعيار آخر لاختيار أفضل نموذج. فإذا كان الانحراف المعياري للنموذج المخفض أقل من أو يساوي الانحراف المعياري للنموذج الكامل، فبالإمكان استخدام النموذج المخفض للحصول على نفس القدرة التنبئية. ويأخذ الانحراف المعياري الصيغة التالية:

$$s = \sqrt{\frac{RSS(p)}{n-p-1}} = \sqrt{MRSS(p)} \qquad (6\text{-}4)$$

والتباين

$$S^2 = \frac{RSS(p)}{n-p-1} = MRSS(p) \qquad (6.5)$$

حيث إن:

s = الانحراف المعياري.

MRSS(p) = متوسط مجموع مربعات البواقي للنموذج المخفض.

RSS(p) = مجموع مربعات البواقي للنموذج المخفض.

p = عدد المتغيرات المستقلة في النموذج المخفض.

n = عدد المشاهدات.

٦-٢-٤ المعيار الرابع - إحصاء ملاوس (Mallows' Cp):

يأخذ إحصاء ملاوس (Cp) الصيغة التالية:

$$Cp = \frac{RSS(p)}{MRSS(k)} - [n - 2(p+1)] \qquad (6\text{-}6)$$

حيث إن:

RSS(p) = مجموع مربعات البواقي للنموذج المخفض.

MRSS(k) = متوسط مجموع مربعات البواقي للنموذج الكامل.

ويساعد إحصاء ملاوس في تحديد عدد المتغيرات المستقلة التي يجب إدخالها في نموذج الانحدار الأفضل، وذلك لأن قيمة إحصاء ملاوس يساوى تقريباً (p+1) إذا كان تباين النموذج المخفض يساوى تقريباً تباين النموذج الكامل (Mallows', 1973; pp 661-675). فالنموذج الجيد هو الذي يساوي عدد معاملاته (عدد المتغيرات +١) قيمة إحصاء ملاوس.

ويستخدم انحراف Cp من عدد معالم النموذج مقياساً للتحيز $(Cp-p-1)$. فالنموذج الذي يكون تحيزه صغيراً، تكون قيمة إحصاء ملاوس قريبةً من عدد معالم النموذج، أي أن:

$$E\left[Cp\middle|Bias=0\right]=\frac{(n-p-1)\sigma^2}{\sigma^2}-n+2(p+1)=p+1 \qquad (6\text{-}7)$$

لذا يقترح ملاوس (Mallows, 1975) أن يتم ترشيح أي نموذج له قيمة ملاوس أقل من عدد معالم النموذج الموفق، أي

$$Cp < (p+1)$$

ويلاحظ مما سبق أن هذه المعايير مرتبطة مع بعضها البعض. فعلى سبيل المثال يمكن الحصول على إحصاء(Fp) الجزئي باستخدام معامل التحديد كما يلي (انظر الفصل الثالث):

$$F_p = \frac{[R^2_{(k)} - R^2_{(p)}]/(k\text{-}p)}{[1\text{-}R^2_{(k)}]/(n\text{-}k\text{-}1)} \sim F_{(k-p),(n-k-1)} \qquad (6\text{-}8)$$

حيث إن:

k = عدد المتغيرات المستقلة في النموذج الكامل، و p = عدد المتغيرات المستقلة في النموذج المخفض.

n = عدد المشاهدات.

$R^2_{(k)}$ = معامل تحديد النموذج الكامل.

$R^2_{(p)}$ = معامل تحديد النموذج المخفض.

وكذلك نجد أن إحصاء ملاوس (Cp) ما هو إلا دالة بسيطة لإحصاء (Fp) حيث

$$Cp = (k\text{-}p)F_p + (2p\text{-}k\text{+}1) \qquad (6\text{-}9)$$

٦-٢-٥ المعيار الخامس - إحصاء مجموع مربعات التنبؤ PRESS:

مجموع مربعات التنبؤ (Prediction Sum of Squares) أو اختصاراً إحصاء PRESS معيار آخر لتقييم القدرة التنبئية لنموذج الانحدار. وإحصاء PRESS هو مجموع مربعات البواقي المحذوفة (انظر الفصل الرابع (Montgomery, Peck and Vining, 2012)، أي:

$$PRESS = \sum_{i=1}^{n}\left(y_i-\hat{y}_{i(i)}\right)^2 = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_{ii}}\right)^2 \qquad (6.10)$$

حيث إن d_i الباقي المحذوف وe_i الباقي العادي، وh_{ii} العنصر القطري رقم i في مصفوفة القبعة $\left(\mathbf{H}=\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\mathbf{X}^{\mathrm{T}}\right)$.

ويلاحظ من المعادلة (6.10) أن قيمة إحصاء PRESS تعتمد على قيمة الباقي العادي وقيمة الرافعة المقابلة له في آن واحد. ويستخدم إحصاء PRESS كمعيار لاختيار النموذج الأفضل؛ فالنموذج الأفضل يجب أن لا يتأثر باستبعاد مشاهدة واحدة. لذا فإن النموذج الأفضل هو الذي لديه أقل قيمة PRESS.

٦-٢-٦ المعيار السادس - معايير المعلومات (SBC, BIC, AIC):

معيار أكايكي للمعلومات Akaike Information Criterion (AIC):

طور أكايكي (Akaike, 1973) معياراً اعتماداً على نظريات المعلومات عرف بمعيار أكايكي للمعلومات للمفاضلة بين نماذج الانحدار الموفقة. ويأخذ المعيار الصيغة التالية (Yan & Su, 2009):

$$\mathrm{AIC}(p)=n\cdot\ln\left(\frac{\mathrm{RSS}(p)}{n}\right)+2\times(p+1) \qquad (6\text{-}11)^{*}$$

ويعد الحد الأول من المعادلة مقياساً لنقص المطابقة للنموذج، ويعد الحد الثاني من المعادلة هو حد جزاء (penalty) لعدد المتغيرات المستقلة. والنموذج الأفضل هو النموذج الذي لديه أقل قيمة لـ AIC.

معيار بايز للمعلومات Bayesian Information Criteria (BIC):

طور ساوا (Sawa, 1978) معيار بايز للمعلومات لاختيار المتغيرات المستقلة في نموذج الانحدار. ويأخذ المعيار الصيغة التالية:

$$\mathrm{BIC}(p)=n\times\ln\left(\frac{\mathrm{RSS}(p)}{n}\right)+\frac{2\times(p+3)\times\mathrm{MRSS}(k)}{\mathrm{RSS}(p)}-\frac{2\times n^{2}\times\mathrm{MRSS}(k)^{2}}{\mathrm{RSS}(p)^{2}} \qquad (6\text{-}12)$$

والنموذج الأفضل هو النموذج الذي لديه أقل قيمة لـ BIC

معيار بايز شوارز Schwarz Bayesian criteria (SBC):

طور شوارز (Schwarz, 1978) معياراً آخر للمفاضلة بين النماذج الموفقة وهو امتداد لمعيار أكايكي للمعلومات. ويأخذ المعيار الصيغة التالية:

---

[*] يستخدم بعض المؤلفين (Greene, 2003; Gujari & Porter, 2009) وبعض برامج الإحصاء (Limdep, Statgraphics) صيغة مختلفة لمعيار **AIC** ، هي: $\left\{\mathrm{AIC}(p)=\ln\left(\frac{\mathrm{RSS}(p)}{n}\right)+\frac{2\times(p+1)}{n}\right\}$

$$SBC(p) = n \times \ln\left(\frac{RSS(p)}{n}\right) + (p+1) \times \ln(n) \qquad (6\text{-}13)^{**}$$

والنموذج الأفضل هو النموذج الذي لديه أقل قيمة لـ SBC.

٦-٢-٧ المعيار السابع - معيار التنبؤ لأمميايا Amemiaya Prediction Criteria:

طور أمميايا (Amemiya, 1980) معياراً للمفاضلة بين النماذج الموفقة. ويأخذ المعيار الصيغة التالية:

$$PC = \frac{(n+p+1)\,MRSS(p)}{TSS} \qquad (6\text{-}14)$$

والنموذج الأفضل هو النموذج الذي لديه أقل قيمة لـ PC.

فيما سبق تم استعراض عدد كبير من معايير المفاضلة، ولكن السؤال لماذا نستخدم أكثر من معيار للمفاضلة بين نماذج الانحدار؟ يرجع السبب في استخدام أكثر من معيار إلى أنه لا يوجد معيار واحد يكون باستمرار هو الأفضل. فمثلاً نجد أن قيمة معامل التحديد تزداد بإضافة أي متغير في حين يمكن أن تنخفض قيمة الخطأ المعياري في حالة إضافة متغير. لذا من الناحية العملية يمكن الحصول على نماذج مختلفة باستخدام هذه المعايير.

٦-٣ طرق اختيار المتغيرات المستقلة في نموذج الانحدار الخطي المتعدد:

٦-٣-١ طريقة اختيار أفضل معادلة من بين كل معادلات الانحدار الممكن توفيقها (All-Possible-Regressions Procedure):

تعتبر هذه الطريقة من أفضل طرق اختيار المتغيرات المستقلة التي تدخل نموذج الانحدار، بل وهي الطريقة الوحيدة التي نضمن بها التوصل إلى نموذج لديه أعلى قيمة لمعامل التحديد (R^2) وإحصاء F وأقل قيمة للخطأ المعياري للتقدير وإحصاء ملاوس(Cp). إلا أن هذه الطريقة يصعب استخدامها باستمرار نظراً لكثرة الحسابات التي تتطلبها، خاصةً إذا كان عدد المتغيرات المستقلة كبيراً.

وتتطلب هذه الطريقة بناء كل النماذج الممكن توفيقها من المتغيرات المستقلة ومن ثم ترتيبها وفقاً للمعايير المذكورة آنفاً لاختيار أفضل نموذج. وبصورة عامة إذا كان لدينا عدد (k) متغير مستقل فإن عدد النماذج الممكن توفيقها يساوي ($\mathbf{2^k - 1}$). فمثلاً إذا كان عدد المتغيرات المستقلة (٤) فإن عدد النماذج الممكن توفيقها يساوي $15 = (2^4 - 1)$

٦-٣-٢ طريقة الحذف إلى الخلف (Backward Elimination Procedure):

في طريقة الحذف إلى الخلف تُتبع الخطوات التالية:

** يستخدم بعض المؤلفين (Greene, 2003; Gujari & Porter, 2009) وبعض برامج الإحصاء (Limdep, Statgraphics) صيغة مختلفة لمعيار **SBC** ، هي: $\left\{SBC(p) = \ln\left(\frac{RSS(p)}{n}\right) + \frac{(p+1)\times\ln(n)}{n}\right\}$

١. يتم أولاً بناء نموذج انحدار يحتوي على جميع المتغيرات المستقلة المرشحة.

٢. يتم حساب إحصاءات F الجزئية لكل متغير على أساس أنه آخر متغير أضيف لبقية المتغيرات المستقلة الأخرى كما يلي:

$$F_1(X_1|X_2,X_3,...,X_k) = \frac{[ESS(X_1,X_2,...,X_k) - ESS(X_1)]/1}{RSS(X_1,X_2,...,X_k)/(n-k-1)}$$

$$F_2(X_2|X_1,X_3,...,X_k) = \frac{[ESS(X_1,X_2,...,X_k) - ESS(X_2)]/1}{RSS(X_1,X_2,...,X_k)/(n-k-1)}$$

$$F_k(X_k|X_1,X_2,...,X_{k-1}) = \frac{[ESS(X_1,X_2,...,X_k) - ESS(X_k)]/1}{RSS(X_1,X_2,...,X_k)/(n-k-1)}$$

٣. يتم تحديد أقل قيمة لإحصاءات F الجزئية ومن ثم يتم إجراء اختبار المعنوية لهذا المتغير، فإذا كانت قيمة إحصاء Fp الجزئي أكبر من قيمة F المستخرجة من جدول توزيع F عند درجتي حرية ١ وn-k-1 ومستوى دلالة محدد (٠,٠٥ مثلاً) فإن ذلك يعني أن جميع المتغيرات المضمنة في النموذج لها تأثير دال إحصائياً على المتغير التابع. وأما إذا كانت قيمة إحصاء F الجزئي أقل من قيمة F المستخرجة من الجدول يتم حذف هذا المتغير من النموذج ويتم بناء نموذج آخر يحتوي على جميع المتغيرات المستقلة عدا المتغير المحذوف.

٤. يتم تكرار الخطوتين (٢) و(٣) إلى أن نصل لنموذج يتضمن متغيرات مستقلة جمعيها ذات دلالة إحصائية.

٦-٣-٣ طريقة الاختيار إلى الأمام (Forward Selection Procedure):

فيما يلي خطوات طريقة الاختيار إلى الأمام لاختيار أفضل نموذج:

١. يتم أولاً حساب معامل الارتباط الخطي البسيط بين المتغير التابع وكل من المتغيرات المستقلة.

٢. يتم بناء نموذج انحدار خطي بسيط بين المتغير التابع والمتغير المستقل الذي لديه أعلى قيمة معامل ارتباط مع المتغير التابع. ومن ثم يتم اختبار معنوية المتغير المستقل باستخدام إحصاء F أو t. فإذا كان المتغير المستقل غير معنوي نخلص إلى أنه لا يوجد متغير مستقل يسهم في تفسير تباين المتغير التابع. وأما إذا كان المتغير المستقل ذا دلالة إحصائية سيتم تضمين هذا المتغير في النموذج وننتقل إلى الخطوة التالية.

٣. في هذه الخطوة يتم بناء نماذج انحدار خطي يضم أي واحد منها متغيرين، أحدهما المتغير الذي تم اختياره في الخطوة الثانية، والآخر من بقية المتغيرات المستقلة، ومن ثم يتم إجراء اختبار F الجزئي للمتغير الذي أضيف للنموذج في هذه الخطوة فقط. فمثلاً إذا كان المتغير الذي تم اختياره في الخطوة الثانية هو المتغير X_1 مثلاً، يتم بناء النماذج التالية:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$$
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_3 + \varepsilon_i$$
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_4 + \varepsilon_i$$
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_{k-1} + \varepsilon_i$$
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_k + \varepsilon_i$$

حيث يلاحظ أن المتغير X_1 متضمن في جميع هذه النماذج.

ومن ثم يتم حساب إحصاءات F الجزئية كما يلي:

$$F_1(X_2|X_1) = \frac{[ESS(X_1X_2) - ESS(X_2)]/1}{RSS(X_1X_2)/(n-3)}$$
$$F_2(X_3|X_1) = \frac{[ESS(X_1X_3) - ESS(X_3)]/1}{RSS(X_1X_3)/(n-3)}$$
..
$$F_k(X_k|X_1) = \frac{[ESS(X_1X_k) - ESS(X_k)]/1}{RSS(X_1X_k)/(n-3)}$$

ومن بعد ذلك يتم اختبار معنوية المتغير الذي لديه أعلى قيمة لإحصاء F الجزئي. فإذا كان المتغير ذا دلالة إحصائية تتم إضافة هذا المتغير لمعادلة الانحدار ليصبح النموذج يضم متغيرين حتى الآن وننتقل إلى الخطوة الرابعة. وأما إذا كان المتغير غير دال إحصائياً، نستخدم نموذج المتغير الواحد الذي اختير في الخطوة السابقة وتتوقف عملية الاختيار.

٤. في هذه الخطوة يتم بناء نماذج تضم ثلاثة متغيرات شريطة أن يضم كل نموذج المتغيرين الذين تمت إضافتهما في المراحل السابقة ويتم إجراء اختبار معنوية المتغير الثالث الذي أضيف في هذه الخطوة بنفس الطريقة التي تم شرحها في الخطوة السابقة. وهكذا في أي خطوة نجد فيها قيمة إحصاء F الجزئي غير معنوي، نعلن أنه لا توجد متغيرات أخرى تسهم في تفسير تباين المتغير التابع وبذلك تنتهي العملية ونعتمد المتغيرات التي تم تضمينها في الخطوة السابقة. وتستمر هذه العملية إلى أن نصل للحد الذي لا يترك فيه أي متغير يمكن أن يسهم في تفسير تباين المتغير التابع. ومن عيوب طريقة الاختيار للأمام، أن أي متغير تم تضمينه في أية خطوة من الخطوات سيظل في النموذج إلى أن تنتهي عملية الاختيار.

٦-٣-٤ طريقة الانحدار التدرجي (Stepwise Regression):

تعتبر طريقة الانحدار التدرجي تطويراً لطريقة الاختيار للإمام وتختلف عنها في أنه يتم اختبار معنوية المتغيرات المستقلة التي سبق إدخالها في نموذج الانحدار في أي خطوة من خطوات عملية الاختيار. وفيما يلي خطوات طريقة الانحدار التدرجي:

١. كما في طريقة الاختيار إلى الأمام، تبدأ هذه الطريقة ببناء نموذج انحدار خطي بسيط يضم المتغير الذي لديه أعلى قيمة معامل ارتباط مع المتغير التابع ومن ثم يتم اختبار معنوية هذا المتغير بالطريقة المعتادة. فإذا كان المتغير المستقل غير معنوي تتوقف عملية الاختيار ونخلص إلى أنه لا يوجد متغير يسهم في تفسير تباين المتغير التابع. وأما إذا كان المتغير المستقل له دلالة إحصائية سيتم ضمه إلى النموذج وننتقل إلى الخطوة الثانية.

٢. في هذه الخطوة يتم اختيار المتغير الثاني الذي سيدخل معادلة الانحدار على أساس قدرته على تفسير أكبر قدر من التباين المتبقي. وبافتراض أن المتغير الذي تم إدخاله في الخطوة الأولي هو X_1، سيتم بناء نماذج انحدار تضم متغيرين أحدهما المتغير الذي اختير في الخطوة الأولى (X_1) وآخر من بقية المتغيرات المستقلة. ومن بعد ذلك يتم حساب إحصاء F الجزئي لكل للمتغيرات التي تم ضمها للمتغير (X_1) لتحديد أعلى قيمة لإحصاء F وبالتالي تحديد المتغير المستقل الثاني الذي سيدخل النموذج. فإذا كان المتغير الذي أُضيف دال إحصائياً يتم تضمينه في النموذج، أما إذا كان غير دال إحصائياً نعلن أنه لا توجد متغيرات أخرى تسهم في تفسير تباين المتغير التابع ويكتفي بنموذج الانحدار الخطي البسيط الذي يضم المتغير (X_1) فقط.

٣. بافتراض أن المتغير الذي أُضيف لنموذج الانحدار في الخطوة الثانية هو (X_3). والآن يكون لدينا نموذج يحتوي على متغيرين هما: (X_1) و(X_3)؛ وقبل البدء في إضافة متغير ثالث يتم اختبار معنوية المتغير الذي أدخل في الخطوة الأولى (X_1) هل ما زال يسهم في تفسير تباين المتغير التابع بمستوى معنوي بوجود المتغير الذي أدخل في الخطوة الثانية (X_3)، أي يتم إجراء الاختبار التالي:

$$F(X_1|X_3) = \frac{\left[ESS(X_1,X_3) - ESS(X_1)\right]/1}{RSS(X_1,X_3)/(n-3)}$$

وفي الخطوات اللاحقة يكون لدينا عدد من إحصاءات F لكل متغير في النموذج بجانب المتغير الذي أُضيف أخيراً، ويكون المتغير الذي لديه أقل قيمة لإحصاء F هو المرشح للحذف. فإذا كانت قيمة أقل إحصاء F أكبر من القيمة الجدولية لتوزيع F سيبقى هذا المتغير في النموذج. وأما إذا كانت قيمة أقل إحصاء F أقل من القيمة الجدولية سيتم حذف هذا المتغير.

٤. نفترض أن المتغير (X_3) ذو دلالة إحصائية وبذلك يكون النموذج يضم المتغيرين (X_1) و(X_3). ولتحديد المتغير الثالث المرشح إدخاله في النموذج يتم إتباع الخطوتين الثانية والثالثة. وتستمر هذه العملية إلى الحد الذي لا يترك فيه أي متغير يمكن أن يسهم في تفسير تباين المتغير التابع.

٦-٤ مثال:

في هذا المثال سيتم تحليل بيانات افتراضية عن المصروفات المعيشية للأسر وبعض المتغيرات المؤثرة عليها (حجم الأسرة، الدخل، عدد الأطفال، مستوى تعليم رب الأسرة) كما يوضح الجدول (٦-١).

جدول رقم (٦-١): بيانات افتراضية عن المصروفات المعيشية للأسرة وبعض المتغيرات المؤثرة عليها

رقم المشاهدة	المصروفات المعيشية (ألف ريال)	مستوى تعليم رب الأسرة (عدد السنوات الدراسية)	عدد الأطفال	دخل الأسرة (ألف ريال)	عدد أفراد الأسرة
1	6.5	6	2	7	5
2	11.2	16	3	13	6
3	11.2	15	4	19	8
4	10.5	14	4	13	8
5	9.3	10	2	9.6	9
6	7.2	10	3	8	5
7	13.4	9	4	15	7
8	5	6	3	5	7
9	11.6	7	4	12	6
10	11.2	12	4	14	7
11	6.2	4	3	7	5
12	12	0	6	19	11
13	11.3	14	6	16	9
14	9.2	5	4	11	8
15	5.5	2	4	6	9
16	6	4	3	7.6	7
17	12.5	16	1	25	8
18	9.8	4	2	10	6
19	6.1	5	3	6.5	5
20	14.3	15	6	15.1	11
21	15.5	15	6	18	10
22	10.8	2	1	11	6
23	4.5	4	1	5.6	5
24	6.7	3	2	8.5	8
25	4.5	2	1	6.3	5
26	9.8	5	3	14	5
27	4	5	1	4.6	7
28	5.5	3	1	7.5	7
29	10	8	2	11.2	7
30	8.5	7	0	9.5	5

يهدف هذا التطبيق إلى توضيح كيفية اختيار "أفضل" نموذج باستخدام طرق الاختيار التي تمت مناقشتها في الجزء (٦-٣) وذلك من خلال مثال يضم أربعة متغيرات مستقلة فقط لسهولة تحليلها وشرحها، إذ كلما زاد عدد المتغيرات المفسرة زاد عدد النماذج الممكن توفيقها، فعلى سبيل المثال يتطلب تحليل عشرة متغيرات بناء ١٠٢٣ نموذجاً.

٦-٤-١ طريقة اختيار أفضل نموذج من بين نماذج الانحدار الممكن توفيقها:

للتوصل إلى أفضل نموذج باستخدام هذه الطريقة، تم أولاً بناء (١٥) نموذجاً خطياً - عدد كل النماذج الممكن توفيقها من (٤) متغيرات مستقلة - ومن ثم تم تلخيص نتائجها في الجدول رقم (٦-٢). فمثلاً بالنسبة للنموذج الأول تم حساب معايير SBC، PC، وBIC، وAIC، وCp، وPRESS كالتالي:

إحصاء SBC:

$$SBC(p) = n \times \ln\left(\frac{RSS(p)}{n}\right) + (p+1) \times \ln(n) = 30 \times \ln\left(\frac{172.81}{30}\right) + (1+1) \times \ln(30) = 59.332$$

إحصاء PC:

$$PC = \frac{(n+p+1)\,MRSS(p)}{TSS} = \frac{(30+1+1) \times 6.172}{291.699} = 0.6771$$

إحصاء BIC :

$$BIC(p) = n \times \ln\left(\frac{RSS(p)}{n}\right) + \frac{2 \times (p+3) \times MRSS(k)}{RSS(p)} - \frac{2 \times n^2 \times MRSS(k)^2}{RSS(p)^2}$$

$$= 30 \times \ln\left(\frac{172.81}{30}\right) + \frac{2 \times (1+3) \times 2.315}{172.81} - \frac{2 \times 30^2 \times 2.315^2}{172.81^2} = 55.422$$

إحصاء AIC :

$$AIC(p) = n \cdot \ln\left(\frac{RSS(p)}{n}\right) + 2 \times (p+1) = 30 \times \ln\left(\frac{172.81}{30}\right) + 2 \times (1+1) = 56.5295$$

إحصاء ملاوس Cp:

باستخدام قيم مجاميع مربعات البواقي الموضحة بالجدول، تم حساب Cp كما يلي:

$$Cp = \frac{RSS(x_1)}{MRSS(x_1,x_2,x_3,x_4)} - [n - 2(p+1)] = \frac{172.81}{2.315} - [30 - 2 \times (1+1)] = 48.65$$

إحصاء PRESS:

$$PRESS = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_{ii}}\right)^2 = \left(\frac{-1.83356}{1-0.036995}\right)^2 + \left(\frac{-1.25713}{1-0.134249}\right)^2 + \ldots + \left(\frac{-0.24592}{1-0.033848}\right)^2 = 201.263$$

وحسب معايير المعلومات AIC وBIC، وPC، وSBC فإن النموذج رقم (٨) هو الأفضل. وأما حسب معياري مقدر الانحراف المعياري (s) ومقدر التباين $\left(s^2\right)$ فإن النموذج الأفضل هو النموذج رقم (١١) يليه النموذج رقم (٨). وحسب معيار إحصائية PRESS فإن النموذج رقم (٦) هو الأفضل يليه النموذج رقم (٨). ويدعم إحصاء ملاوس اختيار النموذج رقم (٨) كأفضل نموذج من بين النماذج الأخرى لأن قيمة (Cp) البالغة (٢,٥) أقل من عدد معالم النموذج وعددها (٣) (الشكل رقم ٦-١).

وبما أن النماذج (١١)، (١٣) و(١٥) لها قيم معاملات تحديد أكبر – وإن كان قليلاً- من قيمة معامل تحديد النموذج رقم (٨) فإنه من الضروري إجراء اختبار ما إذا كان هذا الفرق في القدرة التفسيرية له دلالة إحصائية أم لا؟ ويتطلب هذا الاختبار إجراء ثلاثة اختبارات فرعية هي:

أ) إجراء اختبار دلالة إضافة المتغير (X_1) - تعليم رب الأسرة- على المتغيرين (X_2) و(X_3)، أي هل يسهم المتغير (X_1) في تفسير تباين المتغير التابع في وجود المتغيرين (X_2) و (X_3) ؟. ولإجراء هذا الاختبار نستخدم اختبار F الجزئي، حيث يكون النموذج الكامل هو النموذج رقم (١١) الذي يضم المتغيرات X_1، X_2 و X_3 والنموذج المخفض هو النموذج رقم (٨) الذي يضم المتغيرين(X_2) و (X_3). ويتم اختبار الفرض الصفري الذي نصه "أن متغير تعليم رب الأسرة ليس له تأثير ذو دلالة إحصائية على المصروفات المعيشية" في مقابل الفرض البديل الذي نصه "يؤثر المستوى التعليمي على المصروفات المعيشية للأسر" بواسطة اختبار Fp حيث

$$Fp=\frac{[RSS(p)-RSS(k)]/(k-p)}{RSS(k)/(n-k-1)}=\frac{(61.377-58.071)/1}{58.071/(30-3-1)}=1.48$$

وبما أن قيمة Fp أقل بكثير من القيمة المستخرجة من الجدول $(F_{1,26,\ 0.05}=4.23)$ فإننا لا نستطيع رفض فرض العدم عند مستوى معنوية (٥%). وهذا يعني أن المتغير (X_1) - مستوى تعليم رب الأسرة- لا يسهم في تفسير تباين مصروفات المعيشة.

ب) إجراء اختبار دلالة إضافة المتغير (X_4) - عدد أفراد الأسرة- على المتغيرين (X_2) و(X_3). لإجراء هذا الاختبار نستخدم اختبار F الجزئي، حيث يكون النموذج الكامل هو النموذج رقم (١٣) الذي يضم المتغيرات X_4، X_2 و X_3 والنموذج المخفض هو النموذج رقم (٨) الذي يضم المتغيرين (X_2) و(X_3). وباستخدام اختبار Fp بنفس الطريقة أعلاه نجد أن:

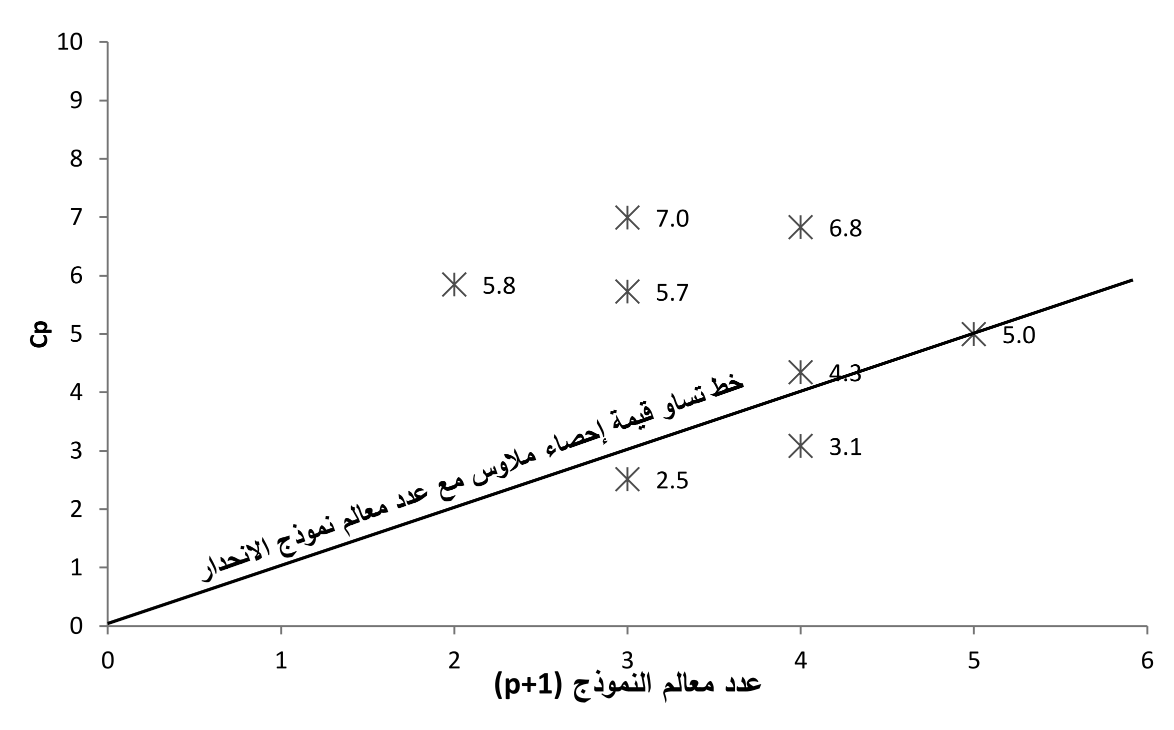
$$Fp=\frac{[RSS(p)-RSS(k)]/(k-p)}{RSS(k)/(n-k-1)}=\frac{(61.377-60.994)/1}{60.994/(30-3-1)}=0.163$$

وبما أن قيمة Fp أقل بكثير من القيمة المستخرجة من الجدول $(F_{1,26,\ 0.05}=4.23)$ فإننا لا نستطيع رفض فرض العدم عند مستوى معنوية (٥%). وهذا يعني أن المتغير (X_4) - عدد أفراد الأسرة - لا يسهم في تفسير تباين مصروفات المعيشة.

ج) وبإتباع نفس طريقة اختبار F الجزئي يمكننا اختبار معنوية إضافة المتغيرين (X_4, X_1) معاً على المتغـيرين (X_2, X_2) حيث يضـم النمـوذج الكامـل - النمـوذج رقـم (١٥)- كـل المتغـيرات في حـين يضـم النمـوذج المخفـض المتغـيرين (X_2, X_3). وبإتباع نفس الطريقة أعلاه يتم حساب إحصاء Fp حيث

$$F_p = \frac{[RSS(p)-RSS(k)]/(k-p)}{RSS(k)/(n-k-1)} = \frac{(61.377-57.874)/2}{57.874/(30-4-1)} = 0.757$$

وبما أن قيمة F الجدولية $(F_{2,25,0.05}=3.39)$ أكبر من قيمة Fp المحسوبة فإننا لا نسـتطيع رفـض فـرض العـدم القائـل بـأن $H_0:\beta_1 = \beta_4 = 0$ بمستوى معنوية ٥%. وهذا يوضح أن المتغيرين (X_4, X_1) معاً لا يؤثران عـلى مصـروفات المعيشـة. ومـن نتائج التحليل أعلاه نجد أن أفضل نموذج هو النموذج رقم (٨). وتوضح نتائج النموذج المستعرضة بالجدول رقم (٦-٢) أن متغيري عدد الأطفال ودخل الأسرة يؤثران على مصـروفات المعيشـة (P-value <0.05) ويفسـر هـذان المتغـيران وحـدهما (٧٩%) من التغير أو التباين الكلي في مصروفات المعيشة.



شكل (٦-١): رسم قيم إحصاء ملاوس مع عدد معالم النماذج التي تقل فيها قيمة إحصاء ملاوس عن (٨).

جدول رقم (٦-٢): ملخص نتائج النماذج الممكن توفيقها

رقم	المتغيرات	RSS	SBC	s	PC	MRSS	BIC	AIC	$\bar{R}^2$	R^2	Cp	PRESS
١	x_1	172.808	59.332	2.484	0.677	6.172	55.422	56.530	0.386	0.408	48.65	201.3
٢	x_2	192.781	62.613	2.624	0.755	6.885	58.433	59.811	0.316	0.339	57.28	222.4
٣	x_3	73.727	33.777	1.623	0.289	2.633	32.736	30.975	0.738	0.747	5.85	100.7
٤	x_4	212.882	65.589	2.757	0.834	7.603	61.183	62.786	0.244	0.270	65.96	239.7
٥	$x_1\ x_2$	132.111	54.677	2.212	0.554	4.893	49.178	50.473	0.514	0.547	33.07	171.4
٦	$x_1\ x_3$	68.821	35.113	1.597	0.288	2.549	32.964	30.909	0.747	0.764	5.73	98.0
٧	$x_1\ x_4$	139.533	56.317	2.273	0.585	5.168	50.595	52.113	0.486	0.522	36.27	176.7
٨	$x_2\ x_3$	61.377	31.679	1.508	0.257	2.273	30.230	27.475	0.774	0.790	2.51	99.8
٩	$x_2\ x_4$	182.974	64.448	2.603	0.767	6.777	57.752	60.244	0.326	0.373	55.04	225.6
١٠	$x_3\ x_4$	71.756	36.366	1.630	0.301	2.658	33.967	32.162	0.736	0.754	7.00	107.0
١١	$x_1\ x_2\ x_3$	58.071	33.419	1.494	0.260	2.234	31.305	27.814	0.778	0.801	3.09	104.6
١٢	$x_1\ x_2\ x_4$	126.530	56.783	2.206	0.567	4.867	49.162	51.178	0.516	0.566	32.66	170.7
١٣	$x_2\ x_3\ x_4$	60.994	34.892	1.532	0.273	2.346	32.358	29.288	0.767	0.791	4.35	104.7
١٤	$x_1\ x_3\ x_4$	66.742	37.594	1.602	0.299	2.567	34.310	31.989	0.745	0.771	6.83	102.7
١٥	$x_1\ x_2\ x_3\ x_4$	57.874	36.718	1.522	0.278	2.315	33.632	29.712	0.770	0.802	5.00	109.3

X_1= مستوى تعليم رب الأسرة (عدد سنوات التعليم)، X_2= عدد الأطفال، X_3=دخل الأسرة (آلاف الريالات)، X_4= عدد أفراد الأسرة

٦-٤-٢ طريقة الحذف إلى الخلف:

لاختيار أفضل نموذج باستخدام طريقة الحذف من الخلف يتم اتباع الخطوات التالية:

- يتم بناء نموذج يضم كل المتغيرات المستقلة – النموذج رقم (١٥) (الجدول ٦-٢).
- يتم حساب إحصاءات F الجزئية لكل متغير على أساس أنه آخر متغير أضيف للنموذج كما يلي:

إحصاء F الجزئي للمتغير X_1:

$$F(X_1\,|\,X_2,X_3,X_4)=\frac{(RSS(X_2,X_3,X_4)-RSS(X_1X_2,X_3,X_4))}{RSS(X_1X_2,X_3,X_4)/(30-4-1)}=\frac{(60.994-57.874)}{57.874/25}=1.35$$

حيث $RSS(X_1,X_2,X_3,X_4)$ هو مجموع مربعات بواقي النموذج الذي يضم جميع المتغيرات و $RSS(X_2,X_3,X_4)$ مجموع مربعات النموذج الذي يضم المتغيرات (X_2,X_3, X_4). وباتباع نفس الطريقة يتم حساب بقية إحصاءات F الجزئية باستخدام قيم مجاميع مربعات البواقي للنماذج الموضحة بالجدول رقم (٦-٢) على النحو التالي:

إحصاء F الجزئي للمتغير X_2:

$$F(X_2|X_1,X_3,X_4) = 3.83$$

إحصاء F الجزئي للمتغير X_3:

$$F(X_3|X_1,X_2,X_4) = 29.66$$

إحصاء F الجزئي للمتغير X_4:

$$F(X_4|X_1,X_2,X_3) = 0.085$$

ويلاحظ أن أقل قيمة لإحصاء F_p هي (٠,٠٨٥) أقل بكثير من قيمة F الجدولية ($F_{1,25,0.05}=4.24$)، ولذلك نقبل فرض العدم، أي أن المتغير X_4 – عدد أفراد الأسرة - لا يسهم في تفسير تباين المتغير التابع بمستوى معنوي (P-value >0.05)[*].

- في هذه الخطوة يتم إسقاط المتغير X_4 وبناء نموذج يضم المتغيرات المتبقية (X_1,X_2,X_3) ويتم اختبار أي متغير على أساس أنه آخر متغير أضيف للنموذج بإتباع نفس الطريقة التي استخدمت في الخطوة السابقة. فيما يلي قيم إحصاءات F الجزئية:

إحصاء F الجزئي للمتغير X_1:

$$F(X_1|X_2,X_3) = 1.48$$

إحصاء F الجزئي للمتغير X_2:

$$F(X_2|X_1,X_3) = 4.81$$

إحصاء F الجزئي للمتغير X_3:

$$F(X_3|X_1,X_2) = 33.14$$

وبما أن أقل قيمة لإحصاء F الجزئي هي (١,٤٨) أقل من قيمة F الجدولية ($F_{1,26,0.05}=4.22$)، فإننا نقبل فرض العدم، أي أن المتغير X_1 (مستوى تعليم رب الأسرة) لا يسهم في تفسير تباين مصروفات المعيشة.

- في هذه الخطوة يتم إسقاط المتغير X_1 ويتم بناء نموذج يضم المتغيرين X_2 وX_3. ويتم إجراء اختبار المتغيرين X_2 وX_3 كل على حدة على أساس إضافة أي منهما للآخر كما يلي:

$$F_p(X_2 | X_3) = \frac{[RSS(X_3) - RSS(X_2,X_3)]}{RSS(X_2,X_3)/(30-2-1)} = \frac{(73.727 - 61.377)}{61.377/27} = 5.433$$

و

$$F_p(X_3 | X_2) = \frac{[RSS(X_2) - RSS(X_2,X_3)]}{RSS(X_2,X_3)/(30-2-1)} = \frac{(192.781 - 61.377)}{61.377/27} = 57.811$$

---

[*] يمكن حساب قيمة الاحتمال باستخدام إكسل، فمثلاً قيمة الاحتمال لاختبار معنوية المتغير المستقل **X_4** نستخدم،
=fdist(0.085;1;25)=0.773

ويلاحظ أن قيمتي Fp المحسوبة للاختبارين أعلاه أكبر من قيمة توزيع F بدرجتي حرية ١ و٢٧($F_{1,27,0.05}$=4.21) وبالتالي نرفض فرضيتي العدم $\beta_2 = 0$ و $\beta_3 = 0$ بمستوى معنوية ٥%، وهذا يعني أن المتغيرين X_2 و X_3 (عدد الأطفال والدخل) يسهمان بمستوى معنوي في تأثيرهما على المتغير التابع.

٦-٤-٣ طريقة الاختيار إلى الأمام:

لاختيار أفضل نموذج باستخدام طريقة الاختيار إلى الأمام يتم اتباع الخطوات التالية:

- تم في الخطوة الأولى في طريقة الاختيار إلى الأمام حساب معاملات الارتباط الخطي البسيط بين المتغير التابع وكل المتغيرات المستقلة المُرشحة. ومن جدول مصفوفة معاملات الارتباط نجد أن أكبر قيمة معامل ارتباط بين المتغير التابع (Y) والمتغير المستقل (X_3) – الدخل - حيث بلغ قيمة معامل الارتباط (r=0.864).

جدول رقم (٦-٣): مصفوفة معاملات الارتباط الخطي البسيط بين المتغيرات المستقلة والمتغير التابع

المتغيرات	X_1	X_2	X_3	X_4
X_2	0.368			
X_3	0.621	0.463		
X_4	0.311	0.655	0.520	
Y	0.638	0.582	0.864	0.520

- في هذه الخطوة تم بناء نموذج انحدار المصروفات المعيشية على الدخل. وتوضح النتائج المستعرضة بالإطار رقم (٦-١) أن الدخل يؤثر بمستوى معنوي (P-value<0.01) على المصروفات المعيشية.

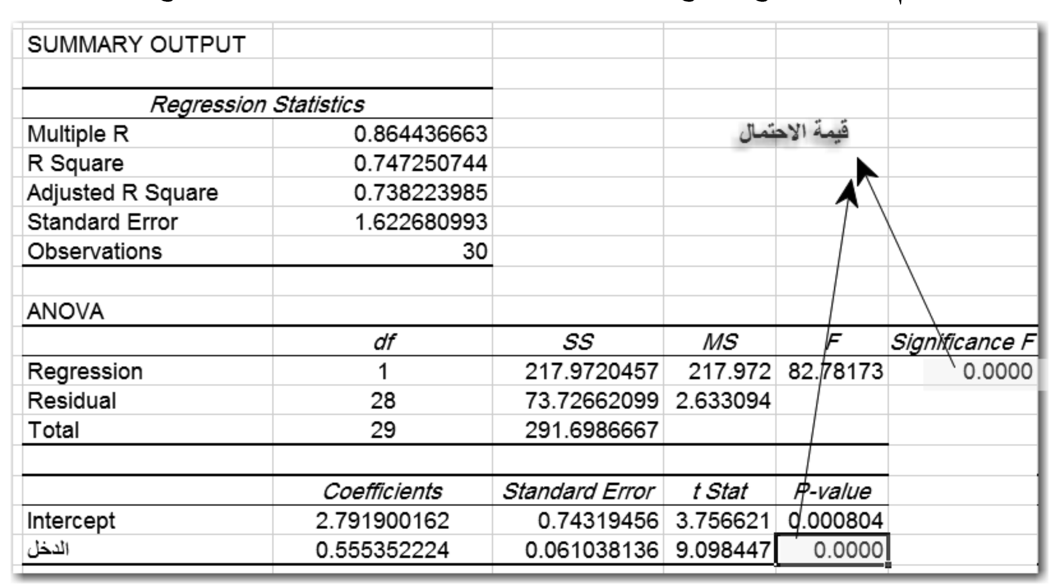
- في هذه الخطوة يتم بناء نماذج ذات متغيرين يضم أي نموذج متغير الدخل الذي دخل النموذج في الخطوة السابقة وأي متغير آخر من بقية المتغيرات المستقلة كما يلي:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + e$$

$$y = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + e$$

$$y = \hat{\beta}_0 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + e$$

إطار رقم (٦-١): نتائج نموذج انحدار المصروفات المعيشية على الدخل (برنامج Excel)



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.864436663
R Square	0.747250744
Adjusted R Square	0.738223985
Standard Error	1.622680993
Observations	30

ANOVA

	df	SS	MS	F	Significance F
Regression	1	217.9720457	217.972	82.78173	0.0000
Residual	28	73.72662099	2.633094		
Total	29	291.6986667			

	Coefficients	Standard Error	t Stat	P-value
Intercept	2.791900162	0.74319456	3.756621	0.000804
الدخل	0.555352224	0.061038136	9.098447	0.0000

ومن نتائج النماذج الثلاثة الموضحة بالجدول رقم (٦-٢) - النماذج رقم (٦) و(٨) و(١٠)- يمكننا حساب قيم F الجزئية لاختبار معنوية إضافة أي متغير للمتغير X_3 على النحو التالي:

لاختبار معنوية إضافة المتغير X_1 يتم اختبار الفرض الصفري $H_0 : \beta_0 \neq 0$ في مقابل الفرض البديل:

$H_1 : \beta_1 \neq 0$ باستخدام F الجزئي التالي:

$$F(X_1 | X_3) = \frac{(RSS(X_3) - RSS(X_1,X_3)/1}{RSS(X_1,X_3)/(30-2-1)} = \frac{(73.727 - 68.821)}{2.549} = 1.925$$

كما يتم اختبار الفرض الصفري $H_0 : \beta_2 = 0$ مقابل الفرض البديل $H_1 : \beta_2 \neq 0$ كما يلي:

$$F(X_2 | X_3) = \frac{(RSS(X_3) - RSS(X_2,X_3)/1}{RSS(X_2,X_3)/(30-2-1)} = \frac{(73.727 - 61.377)}{2.273} = 5.433$$

وكذلك يتم اختبار معنوية إضافة المتغير X_4 كما يلي:

$$F(X_4 | X_3) = \frac{(RSS(X_3) - RSS(X_4,X_3)/1}{RSS(X_4,X_3)/(30-2-1)} = \frac{(73.727 - 71.756)}{2.658} = 0.742$$

ويتضح من نتائج هذه الاختبارات أن إضافة المتغير X_2 (عدد الأطفال) على المتغير X_3 هو المتغير الوحيد الذي يسهم بمستوى معنوي في تفسير تباين مصروفات المعيشة، إذ إن قيمة Fp الجزئي (٥,٤٤) أكبر من قيمة توزيع F بدرجتي حرية ١ و٢٧ عند مستوى معنوية يساوي ٥% ($F_{1,27,0.05}$=4.21).

- في هذه المرحلة يتم بناء نماذج يضم كل واحد ثلاثة متغيرات تتضمن المتغيرين X_2 و X_3 اللذين تم ضمهما في المراحل السابقة كما يلي:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + e$$

$$y = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + e$$

لاختبار معنوية إضافة المتغير X_1 على المتغيرين X_2 و X_3 يتم إجراء اختبار Fp الجزئي كما يلي:

$$F(X_1/X_2,X_3) \quad = \quad 1.48$$

وكذلك يتم حساب Fp الجزئي لاختبار معنوية إضافة المتغير X_4 كما يلي:

$$F(X_4/X_2,X_3) = (61.377-60.994)/2.346 = 0.163$$

وتوضح هذه النتائج أن كلاً من هذين المتغيرين (X_1, X_4) لا يسهم في تفسير تباين المتغير التابع $(P\text{-value}<0.0)$. وعليه نستنتج أن النموذج رقم (٨) حسب الجدول رقم (٦-٢) والذي يضم متغيري الدخل وعدد الأطفال هو النموذج الأفضل من بين النماذج الأخرى.

٦-٤-٤ طريقة الانحدار التدرجي:

- كما في طريقة الاختيار إلى الإمام نبدأ أولاً ببناء نموذج انحدار خطي بسيط يضم المتغير الذي لديه أعلى قيمة معامل ارتباط خطي مع المتغير التابع. والمتغير الذي لديه أعلى قيمة معامل ارتباط مع المتغير التابع هو الدخل، حيث بلغ معامل الارتباط الخطي $(r=0.864)$. وتوضح النتائج المستعرضة بالإطار رقم (٦-١) أن الدخل يؤثر بمستوى معنوي على المصروفات المعيشية $(P\text{-value}<0.05)$.

- في هذه الخطوة يتم بناء نماذج انحدار ذات متغيرين أحدهما المتغير الذي أدخل في الخطوة الأولى (متغير الدخل) كما يلي:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + e$$

$$y = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + e$$

$$y = \hat{\beta}_0 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + e$$

وكما أوضحنا في طريقة الاختيار إلى الأمام نجد أن المتغير X_2 هو المتغير الوحيد الذي يسهم بمستوى معنوي في تفسير تباين مصروفات المعيشة. وبإضافة المتغير X_2 يتضمن النموذج في هذه الخطوة المتغيرين X_3 و X_2. وقبل البدء في إضافة

متغير ثالث يتم اختبار المتغير X_3 أول متغير دخل النموذج. وتعد هذه الخطوة هي الاختلاف الوحيد بين طريقة الاختيار إلى الأمام والاختيار التدرجي. ويتم اختبار معنوية المتغير X_3 في وجود المتغير X_2 كما يلي:

$$F_p(X_3 | X_2) = \frac{[RSS(X_2) - RSS(X_2,X_3)]}{RSS(X_2,X_3)/(30-2-1)} = \frac{(192.781 - 61.377)}{61.377/27} = 57.811$$

وحيث إن قيمة F الجدولية ($F_{1,27,0.05}=4.21$) أقل بكثير من القيمة المحسوبة ($F(X_3|X_2)$)، فإننا نرفض فرض العدم، أي أن الدخل يسهم بمستوى معنوي في تفسير تباين المتغير التابع.

- في هذه الخطوة يتم بناء نماذج ذات ثلاثة متغيرات شريطة أن يتضمن أي نموذج المتغيرين X_3 وX_2. والنماذج الممكن توفيقهما هما:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + e$$
$$y = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + e$$

وتشير النتائج التي تمت مناقشتها في طريقة الاختيار للأمام أن كلاً من X_1 و X_4 لا يسهمان في تفسير تباين المتغير التابع (المصروفات المعيشية) وعليه فإن أفضل نموذج هو النموذج الذي يضم المتغيرين X_3 -الدخل- وX_2 – عدد الأطفال.

٦-٥ اختيار المتغيرات المستقلة باستخدام برنامج نظام التحليل الإحصائي (SAS):

كما سبق شرحه في الفصل الثالث، يتم تحليل البيانات في نظام ساس بإحدى طريقتين، في الطريقة الأولى يتم كتابة برنامج يحدد فيه المتغيرات المراد تحليلها وأداة التحليل الإحصائي من خلال واجهة تطبيق مبنية على نوافذ سهلة الاستخدام، وفي الطريقة الثانية يتم تحليل البيانات من خلال شريط قوائم(Menu-based interface) يتم من خلاله إدخال البيانات أو استيرادها من برامج أخرى كبرنامج إكسل (Excel) وAccess ... الخ أو قراءتها مباشرة من الإنترنت ومن ثم الاختيار من القائمة نوع التحليل المطلوب والحصول مباشرة على نتائج التحليل.

يوضح الإطار رقم (٦-٢) شكل البيانات والأوامر المطلوبة لإجراء تحليل الانحدار باستخدام طريقة الحذف إلى الخلف. فباستخدام الإجراء (Proc reg) يمكن الحصول على نتائج أفضل نموذج باختيار طريقة الاختيار إلى الأمام (Selection=forward) أو اختيار طريقة الانحدار التدرجي (Selection=stepwise).

إطار رقم (٦-٢): شكل بيانات وإجراء Reg وخيار اختيار Backward

```
DATA expenditure;
   INPUT y x1 x2 x3 x4;
     datalines;
6.5      6    2    7        5
11.2     16   3    13       6
11.2     15   4    19       8
.        .    .    .        .
.        .    .    .        .
.        .    .    .        .
8.5      7    0    9.5      5
;
PROC REG corr;
     MODEL y = x1 x2 x3 x4 / selection=backward;
run;
```

ويمكن الحصول على نفس النتائج باستخدام الإجراء (proc stepwise) (الإطار رقم ٦-٣). كما يمكن الحصول على نتائج أفضل نموذج باختيار طريقة الاختيار إلى الأمام بإضافة (model y = x1 x2 x3 x4 /forward) أو اختيار طريقة الانحدار التدرجي (model y = x1 x2 x3 x4 /stepwise).

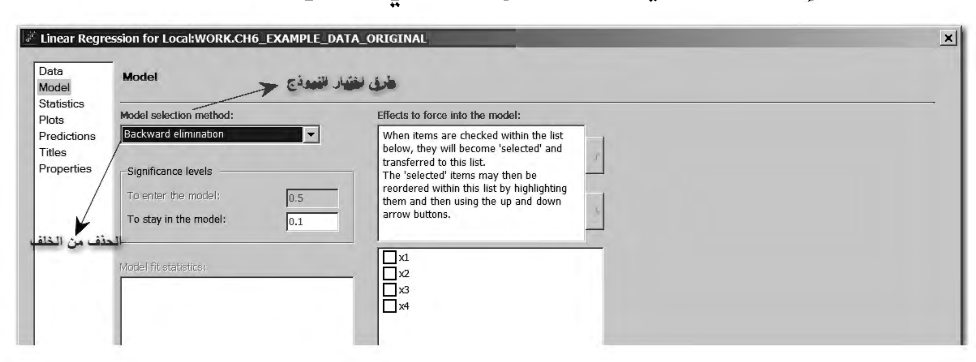
إطار رقم (٦-٣): شكل بيانات وإجراء ساس Stepwise

```
DATA expenditure;
   INPUT y x1 x2 x3 x4;
     datalines;
6.5      6    2    7        5
11.2     16   3    13       6
11.2     15   4    19       8
.        .    .    .        .
.        .    .    .        .
.        .    .    .        .
8.5      7    0    9.5      5
;
PROC stepwise;
     MODEL y = x1 x2 x3 x4 / BACKWARD SLE=0.05 SLS=0.05;
run;
```

وباستخدام برنامج SAS Enterprise يمكن الحصول على نتائج أفضل نموذج باختيار طرق الحذف إلى الخلف والاختيار إلى الأمام والانحدار التدرجي بالإضافة إلى أربعة طرق أخرى*.

* للمزيد يرجى الرجوع **(Freund and Littell, 2000)** و **(Littell, Stroup, and Freund, 2002)**

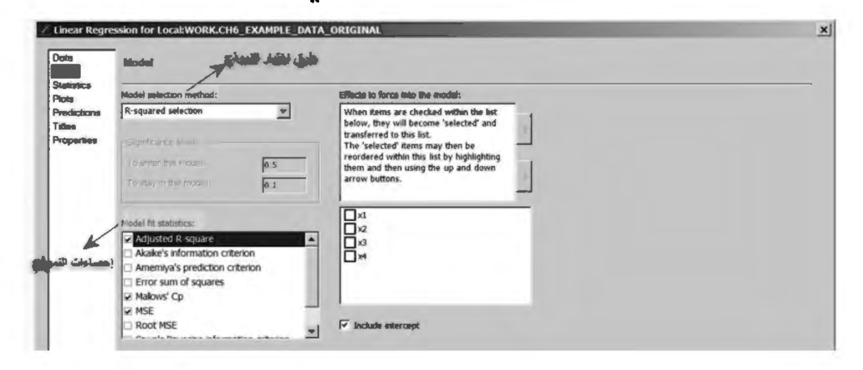
إطار (٦-٤): خيار الحذف من الخلف في SAS Enterprise



٦-٥-١ اختيار أفضل نموذج من بين نماذج الانحدار الممكن توفيقها:

يمكن بناء كل النماذج الممكن توفيقها باستخدام ساس باختيار طريقة R-squared selection (إطار رقم ٦-٥). وباختيار هذه الطريقة يتم بناء جميع النماذج الممكن توفيقها مع جميع الإحصاءات التي تستخدم كمعايير للمفاضلة متوسط مربعات الخطأ MSE، إحصاء ملاوس، ... الخ (الإطار رقم ٦-٤).

إطار (٦-٥): خيار طريقة معامل التحديد في SAS Enterprise



يتضح من نتائج طريقة معامل التحديد (إطار رقم ٦-٦) أن النموذج الذي يضم متغيري الدخل وعدد الأطفال والنموذج الذي يضم مستوى تعليم رب الأسرة وعدد الأطفال ودخل الأسرة هما أفضل النماذج من بين النماذج الأخرى وفقاً لمعايير معامل التحديد المعدل وإحصاء ملاوس ومتوسط مربعات الخطأ. وأخذاً بمعيار إحصاء ملاوس، فإن النموذج الأفضل هو الذي يضم متغيري الدخل وعدد الأطفال لأن قيمة إحصاء ملاوس هي الأقل مقارنة ببقية قيم الإحصاء للنماذج الأخرى، في حين نجد النموذج الذي يضم مستوى تعليم رب الأسرة وعدد الأطفال ودخل الأسرة يعد الأفضل وفقاً لمعياري معامل التحديد المعدل ومتوسط مربعات الخطأ. وبما أن قيمتي معامل التحديد ومتوسط مربعات الخطأ في النموذج الذي يضم عدد الأطفال ودخل الأسرة قريبة من قيمتهما في النموذج الذي يضم مستوى تعليم رب الأسرة وعدد الأطفال ودخل الأسرة، يتم إجراء اختبار معنوية متغير مستوى تعليم رب الأسرة. وفي هذه الحالة يكون النموذج الكامل هو النموذج الذي يضم المتغيرات الثلاثة. وبإجراء اختبارات المعنوية باستخدام اختبار F الجزئي بنفس الطريقة التي سبق

شرحها وجد أن متغير تعليم رب الأسرة ليس لديه تأثير ذو دلالة إحصائية على المصروفات المعيشية. وبالتالي نجد أن النموذج الذي يضم متغيري الدخل وعدد الأطفال هو الأفضل.

إطار (٦-٦): نتائج طريقة اختيار معامل التحديد

R-Square Selection Method

متوسط مربعات الخطأ
احصاء ملاوس
عدد المتغيرات المستقلة
معامل التحديد
معامل التحديد
المتغيرات المستقلة المضمنة في النموذج

Number in Model	R-Square	Adjusted R-Square	C(p)	MSE	Variables in Model
1	0.7473	0.7382	5.8478	2.63309	x3
1	0.4076	0.3864	48.6482	6.17171	x1
1	0.3391	0.3155	57.2759	6.88503	x2
1	0.2702	0.2441	65.9589	7.60291	x4
2	0.7896	0.7740	2.5130	2.27321	x2 x3
2	0.7641	0.7466	5.7286	2.54892	x1 x3
2	0.7540	0.7358	6.9968	2.65765	x3 x4
2	0.5471	0.5135	33.0682	4.89300	x1 x2
2	0.5217	0.4862	36.2742	5.16788	x1 x4
2	0.3727	0.3263	55.0395	6.77680	x2 x4
3	0.8009	0.7780	3.0851	2.23350	x1 x2 x3
3	0.7909	0.7668	4.3479	2.34594	x2 x3 x4
3	0.7712	0.7448	6.8307	2.56700	x1 x3 x4
3	0.5662	0.5162	32.6574	4.86654	x1 x2 x4
4	0.8016	0.7699	5.0000	2.31496	x1 x2 x3 x4

٦-٥-٢ طريقة الحذف إلى الخلف:

يبين الإطار (٦-٧) مخرجات طريقة الحذف إلى الخلف. حيث تتكون المخرجات من أربعة أجزاء هي: (Step 0) وهي الخطوة التي تقابل الخطوة الأولى حسب المثال (الجزء ٦-٤-٢) و(Step 1) وهي تناظر الخطوة الثانية و(Step 2) تناظر الخطوة الثالثة وفي الجزء الأخير يعطى البرنامج ملخصاً لطريقة الاختيار. وتتكون أي خطوة من هذه الخطوات من ثلاثة أجزاء فرعية هي:

- السطر الأول ويحتوي على قيم معامل التحديد (R-square) و قيمة إحصاء ملاوس ((C(p))
- جدول تحليل التباين
- معاملات الانحدار المقدرة ويتكون من ستة أعمدة هي: اسم المتغير (Variable)، قيمة المعامل المقدرة (Parameter Estimate)، الخطأ المعياري (Standard Error)، مجموع مربعات النوع الثاني (Type II Sum of Squares)، قيمة إحصاء F وقيمة الاحتمال (Pr>F). ومجموع مربعات النوع الثاني لأي متغير -(X_j) مثلاً- هو عبارة عن الفرق بين قيمتي مجموع مربعات البواقي للنموذج الذي يضم كل المتغيرات ومجموع مربعات البواقي للنموذج الذي يضم كل المتغيرات باستثناء المتغير (X_j)، أي أن مجموع مربعات النوع الثاني عبارة عن الزيادة في قيمة مجموع مربعات البواقي الناتجة من حذف المتغير المعني من النموذج.

وفيما يلي نستعرض طريقة الحذف من الخلف حسب مخرجات البرنامج:

- في الخطوة الأولى (Step 0) تم بناء نموذج انحدار يضم كل المتغيرات المستقلة المرشحة. ويوضح معامل التحديد أن هذا النموذج يفسر (٨٠,٦%) من التباين في المتغير التابع. وبما أن النموذج الكامل هو نفسه النموذج المخفض في هذه الحالة، فإن قيمة إحصاء ملاوس تكون مساوية لعدد معالم النموذج. ويمكن توضيح ذلك بحساب إحصاء ملاوس حسب الصيغة (٦-٦) كما يلي:

$$Cp = \frac{RSS(k)}{MRSS(k)} - [n - 2(p+1)] = \frac{57.874}{2.315} - [30 - 2 \times (4+1)] \approx 5$$

ويتكون الجزء الثاني من نتائج الخطوة الأولى من جدول تحليل التباين والذي يشير إلى أن الانحدار ككل دال إحصائياً (P-value = < 0.001)، أي أن المتغيرات الأربعة إجمالاً تسهم في تفسير التباين في المتغير التابع (المصروفات المعيشية). وأما الجزء الثاني فيتكون- كما أسلفنا- من اسم المتغير، وقيمة المعلمة المقدرة، والخطأ المعياري، ومجموع مربعات النوع الثاني، وإحصاء F، وقيمة الاحتمال. ولقد تم حساب قيم المعالم المقدرة، والأخطاء المعيارية، وقيمة إحصاء F وقيمة الاحتمال باستخدام طريقة المربعات الصغرى وصيغ الاستدلال الإحصائي كما سبق شرحها في الفصل الثالث. وأما مجموع المربعات من النوع الثاني فقد تم حسابه كما سبق شرحه، فمثلا تم حساب مجموع مربعات النوع الثاني للمعامل الثابت كما يلي:

مجموع مربعات النوع الثاني = مجموع مربعات البواقي للنموذج يحتوي على المتغيرات الأربعة ما عدا المعامل الثابت
ناقصاً
مجموع مربعات البواقي للنموذج الذي يحتوي على المتغيرات المستقلة بما في ذلك المعامل الثابت.
= ٦٨,٠١١١١ - ٥٧,٨٧٤١ = ١٠,١٣٧٠١

ويلاحظ من قيم مجاميع مربعات النوع الثاني أن متغير حجم الأسرة له أقل حجم مساهمة في تفسير تباين المتغير التابع، إذ إن حذفه من النموذج سيؤدي إلى انخفاض مجموع مربعات البواقي بـ ٠,١٩٦٩فقط. ويحتوي العمود الخامس على قيم إحصاءF ويتم حسابها حسب الصيغ التي تم شرحها في الفصل الثالث. فمثلاً بالنسبة للمعامل الثابت يتم حساب إحصاء F بإحدى الطريقتين التاليين:

حساب مربع إحصاء T:

$$F_p = t^2 = \left(\frac{b_0}{s.e(b_o)}\right)^2 = \left(\frac{2.50049}{1.19493}\right)^2 = 4.38$$

أو بحساب إحصاء F الجزئي:

$$F_p = \frac{RSS(X_1,X_2,X_3,X_4) - RSS(X_0,X_1,X_2,X_3,X_4)}{RSS(X_0,X_1,X_2,X_3,X_4)/(n-p-1)} = \frac{(68.01111 - 57.841)}{57.841/25} \simeq 4.40$$

أما القيمة (Prob>F) فهي قيمة احتمال توزيع F المناظرة لقيمة معينة (X) وبدرجتي حرية واحد صحيح و $(n-p-1)$. فمثلاً بالنسبة للمعامل الثابت نجد أن قيمة احتمال توزيع F المناظرة للقيمة (٤,٣٨) بدرجتي حرية (١)

و(٢٥) هي (٠,٠٤٦٧). حيث يتم استخراج قيمة الاحتمال إما باستخدام جدول توزيع F (انظر ملحق الجداول الإحصائية) أو باستخدام الحاسب الآلي (برنامج SAS، SPSS، MINITAB و STATA وEXCEL).

وتوضح نتائج الخطوة الأولى أن متغير حجم الأسرة له أقل حجم مساهمة في تفسير تباين المتغير التابع، وأن هذه المساهمة غير دالة إحصائياً(P-value=0.44). وبذلك يكون هذا المتغير هو المرشح الأول للخروج من النموذج.

وفي الخطوة الثانية تم حذف متغير حجم الأسرة وبناء نموذج من المتغيرات الثلاثة المتبقية - عدد الاطفال، تعليم رب الأسرة والدخل - ويفسر هذا النموذج نحو (٨٠,١%) من تباين المتغير التابع وبلغت قيمة إحصاء ملاوس (٣,١) ويوضح جدول تحليل التباين أن هناك علاقة خطية معنوية بين المتغير التابع والمتغيرات المستقلة المضمنة في النموذج. وبالنظر إلى عمودي "F" و "Prob>F" نجد أن متغير تعليم رب الأسرة له أقل قيمة لإحصاء F وغير دال إحصائياً (P-value =0.235). وبذلك يكون هذا المتغير هو المرشح للخروج من النموذج.

وفي الخطوة الثالثة تم حذف متغير تعليم رب الأسرة وتم بناء نموذج من المتغيرين المتبقيين - الدخل وعدد الأطفال - ويفسر هذا النموذج نحو (٧٩%) من تباين المتغير التابع وبلغت قيمة إحصاء ملاوس (٢,٥). كما يوضح جدول تحليل التباين أن الانحدار ككل دال إحصائياً. وتوضح قيم إحصاء F وقيم الاحتمال أن متغيري الدخل وعدد الأطفال لهما دلالة إحصائية في تأثيرهما على المتغير التابع. وبعد هذه الخطوة يعلن النظام أن كل المتغيرات المضمنة في النموذج دالة إحصائياً عند مستوى معنوية (٥%). وتشمل المخرجات ملخصاً لطريقة الحذف من الخلف يشتمل على معلومات عن المتغيرات التي تم حذفها في كل خطوة. حيث يوضح العمود الثاني عدد المتغيرات المضمنة في النموذج في الخطوة الثانية ويوضح العمود الثالث - معامل التحديد الجزئي (Partial R-square)- الانخفاض في قيمة معامل التحديد عند حذف هذا المتغير، والعمود الرابع - معامل تحديد النموذج Model R-square- يوضح قيمة معامل التحديد بعد حذف المتغير، ويوضح العمود الخامس قيمة إحصاء ملاوس "Cp" بعد حذف المتغير، ويوضح العمود قيمة إحصاء "F" المناظر للمتغير قبل الحذف والعمود "Prob>F" قيمة الاحتمال المناظرة للمتغير قبل الحذف. فمثلاً قام البرنامج بحساب إحصاءات متغير حجم الأسرة (FSZE) المحذوف في الخطوة الأولى كما يلي:

معامل التحديد الجزئي = معامل التحديد لنموذج الانحدار الذي يضم كل المتغيرات ناقصاً معامل التحديد للنموذج الذي يضم كل المتغيرات ماعدا حجم الأسرة

معامل التحديد الجزئي = ٠,٨٠١٦ - ٠,٨٠٠٩ = ٠,٠٠٠٧

معامل تحديد النموذج = معامل التحديد لنموذج الانحدار الذي يضم كل المتغيرات ناقصاً مساهمة المتغير المعني في معامل التحديد

معامل تحديد النموذج = ٠,٨٠١٦ - ٠,٠٠٠٧ = ٠,٨٠٠٩

C(p)= قيمة إحصاء ملاوس بعد حذف متغير حجم الأسرة (انظر الخطوة Step 1 من المخرجات).

F = قيمة إحصاء F المناظرة للمتغير قبل الحذف (انظر الخطوة Step 0 من المخرجات).

Prob>F= قيمة الاحتمال المناظرة للمتغير قبل الحذف (انظر الخطوة Step 0 من المخرجات).

إطار رقم (٦-٧-أ): نتائج طريقة الحذف إلى الخلف

Backward Elimination: Step 0

All Variables Entered: R-Square = 0.8016 and C(p) = 5.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	233.82457	58.45614	25.25	<.0001
Error	25	57.87410	2.31496		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.50049	1.19493	10.13701	4.38	0.0467
x1	0.08622	0.07426	3.12029	1.35	0.2566
x2	0.45094	0.23040	8.86786	3.83	0.0616
x3	0.44369	0.08147	68.65581	29.66	<.0001
x4	-0.06435	0.22064	0.19690	0.09	0.7730

إطار رقم (٦-٧-ب): نتائج طريقة الحذف إلى الخلف

Backward Elimination: Step 1

Variable x4 Removed: R-Square = 0.8009 and C(p) = 3.0851

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	233.62767	77.87589	34.87	<.0001
Error	26	58.07100	2.23350		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.22571	0.72193	21.22915	9.50	0.0048
x1	0.08832	0.07260	3.30565	1.48	0.2347
x2	0.41381	0.18862	10.74974	4.81	0.0374
x3	0.43601	0.07573	74.03990	33.15	<.0001

إطار رقم (٦-٧-ج): نتائج طريقة الحذف إلى الخلف

Backward Elimination: Step 2

Variable x1 Removed: R-Square = 0.7896 and C(p) = 2.5130

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	230.32201	115.16101	50.66	<.0001
Error	27	61.37665	2.27321		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.25528	0.72791	21.82167	9.60	0.0045
x2	0.44053	0.18900	12.34997	5.43	0.0275
x3	0.48637	0.06397	131.40406	57.81	<.0001

إطار رقم (٦-٧-د): نتائج طريقة الحذف إلى الخلف

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4	3	0.0007	0.8009	3.0851	0.09	0.7730
2	x1	2	0.0113	0.7896	2.5130	1.48	0.2347

٦-٥-٣ طريقة الاختيار إلى الأمام:

يوضح الإطار رقم (٦-٨) مخرجات نظام ساس لطريقة الاختيار للأمام. حيث يلاحظ أن شكل المخرجات متماثل تماماً لمخرجات طريقة الحذف من الخلف. وكما سبق شرحه تبدأ الطريقة ببناء انحدار خطي بسيط مع المتغير الذي لديه أعلى قيمة ارتباط خطي مع المتغير التابع. ففي الخطوة الأولى (Step 1) تم إدخال متغير الدخل ذلك لأن لهذا المتغير أعلى قيمة معامل ارتباط مع المتغير التابع. وتوضح النتائج أن الانحدار ككل دال إحصائياً (P-value<0.0001). وفي الخطوة الثانية تم إدخال متغير عدد الأطفال في النموذج على أساس أنه المتغير الوحيد الذي يسهم بمستوى معنوي في تفسير تباين المتغير التابع. ومن بعد ذلك يعلن النظام أنه لا توجد متغيرات أخرى تسهم في تفسير المتغير التابع عند مستوى معنوية ٥%. كما يعطي النظام ملخصاً لعملية الاختيار للأمام مماثل لملخص طريقة الحذف من الخلف. حيث يتم تلخيص لإحصاءات خطوات الاختيار والتي تشمل عدد المتغيرات، معامل التحديد، قيمة إحصاء ملاوس، إحصاء F وقيمة الاحتمال. ففي الخطوة الأولى تم تلخيص لإحصاءات النموذج الخطي البسيط وفي الثانية تم توضيح التغير الذي طرأ على النموذج بعد إدخال المتغير الثاني فمثلاً أصبح عدد المتغيرات اثنين وبلغت قيمة المساهمة الخاصة بالمتغير الثاني (عدد الأطفال) في معامل التحديد (٠,٠٤٢٣) وبذلك تزيد قيمة معامل التحديد من (٠,٧٤٧٣) في نموذج الانحدار البسيط لتصبح (٠,٧٨٩٦ = ٠,٧٤٧٣+٠,٠٤٢٣) وتنخفض قيمة إحصاء ملاوس إلى (٢,٥١٣).

إطار رقم (٦-٨-أ): نتائج طريقة الاختيار إلى الأمام

Forward Selection: Step 1

Variable x3 Entered: R-Square = 0.7473 and C(p) = 5.8478

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	217.97205	217.97205	82.78	<.0001
Error	28	73.72662	2.63309		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.79190	0.74319	37.15874	14.11	0.0008
x3	0.55535	0.06104	217.97205	82.78	<.0001

إطار رقم (٦-٨-ب): نتائج طريقة الاختيار إلى الأمام

Forward Selection: Step 2

Variable x2 Entered: R-Square = 0.7896 and C(p) = 2.5130

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	230.32201	115.16101	50.66	<.0001
Error	27	61.37665	2.27321		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.25528	0.72791	21.82167	9.60	0.0045
x2	0.44053	0.18900	12.34997	5.43	0.0275
x3	0.48637	0.06397	131.40406	57.81	<.0001

إطار رقم (٦-٨-ج): نتائج طريقة الاختيار إلى الأمام

Summary of Forward Selection							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3	1	0.7473	0.7473	5.8478	82.78	<.0001
2	x2	2	0.0423	0.7896	2.5130	5.43	0.0275

٦-٥-٤ طريقة الانحدار التدرجي:

يوضح الإطار رقم (٦-٩) مخرجات تحليل الانحدار التدرجي. ويلاحظ من النتائج أن الخطوة الأولى في طريقة الانحدار التدرجي هي نفس الخطوة الأولى لطريقة الإضافة للأمام. أما في الخطوة الثانية فوجد أن من بين النماذج التي تضم في

كل منها متغيرين، أن النموذج الذي يضم متغيري الدخل وعدد الأطفال هو أفضل النماذج. أما بقية نتـائج الطريقـة فهـي متطابقة تماماً مع نتائج طريقة الاختيار إلى الأمام ولها نفس التفسير.

إطار رقم (٦-٩-أ): نتائج طريقة الانحدار التدرجي

Stepwise Selection: Step 1

Variable x3 Entered: R-Square = 0.7473 and C(p) = 5.8478

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	217.97205	217.97205	82.78	<.0001
Error	28	73.72662	2.63309		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.79190	0.74319	37.15874	14.11	0.0008
x3	0.55535	0.06104	217.97205	82.78	<.0001

إطار رقم (٦-٩-ب): نتائج طريقة الانحدار التدرجي

Stepwise Selection: Step 2

Variable x2 Entered: R-Square = 0.7896 and C(p) = 2.5130

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	230.32201	115.16101	50.66	<.0001
Error	27	61.37665	2.27321		
Corrected Total	29	291.69867			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.25528	0.72791	21.82167	9.60	0.0045
x2	0.44053	0.18900	12.34997	5.43	0.0275
x3	0.48637	0.06397	131.40406	57.81	<.0001

إطار رقم (٦-٩-ج): نتائج طريقة الانحدار التدرجي

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3		1	0.7473	0.7473	5.8478	82.78	<.0001
2	x2		2	0.0423	0.7896	2.5130	5.43	0.0275

٦-٦ ملاحظات:

على الرغم من أهداف ومزايا طرق اختيار أفضل نموذج انحدار والتي من أهمها تقليل عدد معالم نموذج الانحدار الخطي المتعدد، إلا أن هناك عدداً من الملاحظات التي يجب مراعاتها عند استخدامها.

- تقوم نظرية هذه الطرق على التحديد الصحيح للصيغة الدالية بين المتغير التابع والمتغيرات المستقلة.
- تفترض هذه الطرق أن المتغيرات التابعة والمستقلة لا تتضمن بيانات شاذة أو مؤثرة.
- يمكن الحصول على نماذج انحدار "أفضل" مختلفة تبعاً للطريقة المستخدمة. فمثلاً قد نحصل باستخدام الانحدار التدرجي على نموذج "أفضل" يختلف عن النموذج الذي يمكن الحصول عليه باستخدام طريقة الحذف إلى الخلف.
- استخدام هذه لا يؤدي دائماً إلى الوصول لأفضل نموذج؛ إذ يمكن الوصول إلى نموذج أفضل بدون استخدام هذه الطرق.
- إن طرق اختيار المتغيرات المستقلة التي تمت مناقشتها في هذا الفصل يجب استخدامها بحذر، إذ إن الاختيار الآلي للمتغيرات يجب أن لا يفسر كانعكاس للأهمية النسبية لهذه المتغيرات.
- توجد أحياناً أسباب نظرية تحتم الاحتفاظ ببعض المتغيرات في النموذج. ولحسن الحظ تتيح بعض برامج الإحصاء الجاهزة خيار تثبيت بعض المتغيرات في النموذج وترك البعض الآخر للدخول أو الخروج آلياً.
- لا تستخدم هذه الطرق عند الاختيار أي فحوص تشخيصية للنموذج كتحليل البواقي، ولذلك من المحتمل أن نحصل على نموذج "أفضل" لكنه غير مستوف لاشتراطات نموذج الانحدار.
- لا تتيح هذه الطرق إجراء تحويلات للمتغير التابع أو لمتغير أو أكثر من المتغيرات المستقلة في الحالات التي تتطلب ذلك.
- يجب استخدام هذه الطرق بحذر خاصة في وجود متغيرات تفاعل (Interaction repressors) ضمن المتغيرات المستقلة؛ إذ من الممكن اختيار متغير تفاعل دون اختيار المكونات الأساسية المكونة له مما ينافي مبدأ التهميش (Principle of Marginality).
- إن التطور الكبير في مجال الحاسب جعل عملية بناء كل النماذج الممكن توفيقها سهلة وسريعة حتى في حالة وجود عدد كبير من المتغيرات المستقلة. وتعتبر طريقة اختيار أفضل نموذج من بين النماذج الممكن توفيقها أفضل طريقة للاختيار مقارنة بالطرق الأخرى.

تمارين

١. وضح أوجه الشبه والاختلاف بين طريقتي الإضافة إلى الأمام والانحدار التدرجي.

٢. البيانات التالية توضح بعض نتائج نماذج انحدار المتغير التابع (Y) على ثلاثة متغيرات (X_1, X_2, X_3) :

المتغيرات المفسرة (Explanatory Variables)	مجموع مربعات البواقي (Residual Sum of Squares)
X_1	77.6
X_2	102.2
X_3	76.0
X_1, X_2	30.5
X_1, X_3	32.6
X_2, X_3	56.8
X_1, X_2, X_3	25.0

فإذا كان عدد المشاهدات التي استخدمت في بناء هذه النماذج يساوي (٢٠) ومجموع المربعات الكلي (TSS) يساوي ١٥٦,٧، استخدم طرق اختيار الحذف إلى الخلف، الاختيار إلى الأمام والاختيار التدرجي للحصول على أفضل نموذج عند مستوى معنوية ٥%. أوجد قيم معامل التحديد للنماذج المختارة.

الفصل السابع

مشكلات عدم استيفاء اشتراطات
طريقة المربعات الصغرى الاعتيادية وطرق معالجتها

تناولنا في الفصلين الثاني والثالث اشتراطات استخدام طريقة المربعات الصغرى في تقدير معالم نموذج الانحدار. وتشمل هذه الاشتراطات، اشتراطات خاصة بحد الخطأ العشوائي، وبعضها يخص المتغيرات المستقلة والبعض الآخر يخص العلاقة بين حد الخطأ العشوائي والمتغيرات المستقلة. وباستيفاء هذه الاشتراطات تتصف مقدرات المربعات الصغرى لمعالم نموذج الانحدار الخطي بأنها أفضل تقدير خطي غير متحيز. وفي حالة عدم استيفاء بعض هذه الاشتراطات نواجه بمشاكل تعرف بمشاكل القياس التي لا بد من معالجتها للحصول على مقدرات المربعات الصغرى ذات الخصائص الحميدة المتمثلة في عدم التحيز والاتساق والكفاءة. سنتناول في هذا الفصل أهم أربع مشكلات، هي: أخطاء توصيف نموذج الانحدار، والارتباط الخطي المتعدد، واختلاف التباين، والارتباط الذاتي.

٧-١ أخطاء توصيف نموذج الانحدار الخطي (Model Misspecification):

تعد مرحلة توصيف النموذج من أهم مراحل بناء نموذج الانحدار؛ إذ إن أي خطأ في هذه المرحلة قد يسهم في الوصول إلى نتائج خاطئة. ويتضمن توصيف نموذج الانحدار الخطي ما يلي:

١. التحديد الصحيح للصيغة الدالية بين المتغير التابع والمتغيرات المستقلة. وكما أشرنا في الفصل الثاني، يُعد شكل الانتشار الأداة الأساسية لتحديد شكل العلاقة بين المتغير التابع والمتغيرات المستقلة. فمن شكل الانتشار يمكن تحديد ما إذا كانت العلاقة بين المتغير التابع وأي من المتغيرات المستقلة خطية أو غير خطية ربما تحتاج إلى إجراء تحويلة أو تحويلات في المتغير التابع أو المتغيرات المستقلة، أو كليهما معاً. بالإضافة إلى ذلك يتم في هذه المرحلة الاعتماد على النظريات العلمية والدراسات والبحوث التطبيقية لتحديد شكل العلاقة الدالية واتجاه العلاقة بين المتغير التابع وكل من المتغيرات المستقلة. فمثلاً في علم الاقتصاد يُستخدم نموذج دالة الإنتاج لكوب دوجلاس (Cobb Douglas) لقياس العلاقة بين الإنتاج (Y) كمتغير تابع، وكل من رأس المال (C) والعمالة (L) باستخدام العلاقة الدالية: $Y = \beta_0 C^{\beta_1} L^{\beta_2} e^{\varepsilon}$.

٢. بالإضافة إلى الصياغة الدالية الصحيحة، هناك اعتبارات أخرى لا بد من أخذها في الحسبان، من أهمها أن يكون النموذج المراد بناؤه خطي المعالم (Linear in parameters) أو يمكن تحويله إلى نموذج خطي المعالم، وأن تكون الصيغة الدالية مبسطة قدر الإمكان اتساقاً مع مبدأ التبسيط أو الاختصار (Parsimony principle)، ويقصد بالمبدأ أن يتم اختيار النموذج الأبسط الذي يحتوي على المتغيرات المستقلة ذات العلاقة ذات الصياغة الدالية غير المعقدة فقط ؛ إذ إن النموذج الأبسط سهل الفهم والاستخدام.

٣. إدراج جميع المتغيرات المستقلة ذات العلاقة في نموذج الانحدار الخطي. فإسقاط متغيرات مستقلة ذات علاقة في النموذج أو إدخال متغيرات غير ذات علاقة في النموذج من أهم المشكلات التي تواجه الباحث في هذه المرحلة. ومن أسباب عدم تضمين النموذج للمتغيرات المستقلة ذات العلاقة عدم توفر بيانات كافية عنها، وعدم التعرف على تلك المتغيرات، واعتماد الباحث على نظريات تم تطويرها وتطبيقها في مجتمعات مختلفة عن مجتمع الدراسة الذي جمع منه البيانات. وأما إدخال متغيرات مستقلة غير ذات علاقة تحدث أحياناً في بعض البحوث المشاهدة (Observational Research) حيث لا يستطيع الباحث معرفة جميع المتغيرات المستقلة المفسرة للمتغير التابع.

وتسمى أخطاء الصياغة الدالية وإسـقاط متغـيرات مسـتقلة ذات علاقـة أو إدخـال متغـيرات مسـتقلة ليسـت ذات علاقة بأخطاء توصيف نموذج الانحدار الخطي. فمشكلة الصياغة الدالية تم معالجتها في الفصول الثلاثـة الأولى. وتعـد أخطاء التوصيف المتعلقة بعدم تضمين النموذج متغيرات مستقلة ذات علاقة وإدخال متغيرات مستقلة غير ذات علاقة من أكثر مشاكل أخطاء التوصيف حدوثاً (Greene, 2006, p.148). لذا سنتناول في هذا الجزء مشكلتي إسـقاط متغـيرات مستقلة ذات علاقة وإدخال متغيرات مستقلة ليست ذات علاقة في نموذج الانحدار الخطي.

٧-١-١ عدم إدخال المتغيرات المستقلة ذات العلاقة (Omission of Relevant Variables):

يقصد بعدم إدخال متغيرات مستقلة ذات علاقة أن نموذج الانحدار الموفق لم يتضـمن متغـيراً أو متغـيرات مسـتقلة ذات علاقة تسهم في تفسير تباين المتغير التابع. وعدم إدخال المتغـيرات المسـتقلة ذات العلاقـة في النمـوذج يـؤدي إلى الحصول على مقدرات متحيزة لكل من معاملات الانحدار وتقدير التباين والإحصـاءات الأخـرى المرتبطـة بهـما. وفـيما يلي نوضح رياضياً أثر عدم إدخال المتغيرات ذات العلاقة على تقدير معاملات نموذج الانحدار وتقدير التباين.

تقدير معالم نموذج الانحدار الخطي في ظل إسقاط متغيرات مستقلة ذات علاقة:

بفرض أن لدينا $(p+r)$ متغيرات مستقلة جمعيها ذات علاقة، وتم بناء نموذج انحدار خطي من r متغيرات مستقلة بإسقاط (p) متغيرات مستقلة، يمكن تجزئة مصفوفة نموذج الانحدار الخطي المتعدد كالتالي:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \boldsymbol{\varepsilon} \qquad (7\text{-}1)$$

حيث إن:

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{1}\ X_1\ X_2\ \ldots\ X_r \end{bmatrix}$$
$$\mathbf{X}_2 = \begin{bmatrix} X_{r+1}\ X_{r+2}\ \ldots\ X_p \end{bmatrix}$$
$$\mathbf{X} = \begin{bmatrix} 1\ X_1\ X_2\ \ldots X_r \vdots X_{r+1} \ldots X_p \end{bmatrix} = [\mathbf{X}_1\ \mathbf{X}_2]$$

إذا ما تم بناء نموذج انحدار خطي من r متغيرات مستقلة، $\mathbf{X}_1$، بإسقاط (p) متغيرات مستقلة $\mathbf{X}_2$، نحصل على مقدرات معالم كما يلي:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1\right)^{-1} \mathbf{x}_1^{\mathrm{T}}\mathbf{y}$$

وبإيجاد التوقع، نحصل على

$$\begin{aligned} E\left(\widehat{\boldsymbol{\beta}}_1\right) &= \left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}\boldsymbol{\beta} \\ &= \left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\left(\mathbf{x}_1^T\mathbf{x}_1 \quad \mathbf{x}_1^T\mathbf{x}_2\right)\boldsymbol{\beta} \\ &= \left(\mathbf{I} \quad \left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\right)\begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix} \\ &= \boldsymbol{\beta}_1 + \left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\boldsymbol{\beta}_2 \end{aligned} \quad (7\text{-}2)$$

ويلاحظ من النتيجة أن مقدرات معالم النموذج $\widehat{\boldsymbol{\beta}}_1$ متحيزة للمعالم $\boldsymbol{\beta}_1$، ويمكن أن تكون هذه المقدرات غير متحيزة في حالتين فقط، هما: أن تكون قيم معالم المتغيرات المستقلة المسقطة مساوية للصفر $\boldsymbol{\beta}_2 = \mathbf{0}$ أو أن تكون المتغيرات المدخلة متعامدة (Orthogonal) مع المتغيرات غير المدخلة في النموذج ، أي $\mathbf{x}_1^T\mathbf{x}_2 = \mathbf{0}$. وبالتالي لم تعد مقدرات النموذج في ظل عدم إدخال متغيرات ذات علاقة تتميز بأنها أفضل مقدرات خطية غير متحيزة.

تقدير التباين في ظل إسقاط متغيرات مستقلة ذات علاقة:

نعلم أن مجموع مربعات البواقي:

$$\begin{aligned} RSS_1 &= \mathbf{e}^T\mathbf{e} \\ &= \left(\mathbf{M}_1\mathbf{Y}\right)^T\mathbf{M}_1\mathbf{Y} \\ &= \mathbf{Y}^T\mathbf{M}_1\mathbf{Y} \\ &= \left(\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)^T\mathbf{M}_1\left(\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right) \\ &= \left(\boldsymbol{\beta}^T\mathbf{x}^T\mathbf{M}_1+\boldsymbol{\varepsilon}^T\mathbf{M}_1\right)\left(\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right) \\ &= \boldsymbol{\beta}^T\mathbf{x}^T\mathbf{M}_1\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\beta}^T\mathbf{x}^T\mathbf{M}_1\boldsymbol{\varepsilon}+\boldsymbol{\varepsilon}^T\mathbf{M}_1\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon}^T\mathbf{M}_1\boldsymbol{\varepsilon} \end{aligned}$$

حيث إن $\mathbf{M}_1 = \mathbf{I}-\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T$

وللحصول على مقدر التباين، نحسب القيمة المتوقعة لمجموع مربعات البواقي كما يلي(Ismail,2012):

$$\begin{aligned} E\left(RSS_1\right) &= E\left(\boldsymbol{\beta}^T\mathbf{x}^T\mathbf{M}_1\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\beta}^T\mathbf{x}^T\mathbf{M}_1\boldsymbol{\varepsilon}+\boldsymbol{\varepsilon}^T\mathbf{M}_1\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon}^T\mathbf{M}_1\boldsymbol{\varepsilon}\right) \\ &= \sigma^2 \text{trace}\left(\mathbf{I}-\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\right)+\left(\boldsymbol{\beta}_1^T \quad \boldsymbol{\beta}_2^T\right)\begin{pmatrix}\mathbf{x}_1^T \\ \mathbf{x}_2^T\end{pmatrix}\left(\mathbf{I}-\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\right)\left(\mathbf{x}_1 \quad \mathbf{x}_2\right)\begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix} \end{aligned}$$

$$E\left(RSS_1\right)=\sigma^2\text{trace}\left(\mathbf{I}-\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\right)+\begin{pmatrix}\boldsymbol{\beta}_1^T & \boldsymbol{\beta}_2^T\end{pmatrix}\begin{pmatrix}\mathbf{x}_1^T\\ \mathbf{x}_2^T\end{pmatrix}\left(\mathbf{I}-\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\right)\begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$$

$$=\sigma^2\text{trace}\left(\mathbf{I}-\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\right)+\begin{pmatrix}\boldsymbol{\beta}_1^T & \boldsymbol{\beta}_2^T\end{pmatrix}\begin{pmatrix}\mathbf{x}_1^T-\mathbf{x}_1^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\\ \mathbf{x}_2^T-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\end{pmatrix}\begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$$

$$E\left(RSS_1\right)=\left(n-r-1\right)\sigma^2+\begin{pmatrix}\boldsymbol{\beta}_1^T & \boldsymbol{\beta}_2^T\end{pmatrix}\begin{pmatrix}0\\ \mathbf{x}_2^T-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\end{pmatrix}\begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$$

$$=\left(n-r-1\right)\sigma^2+\begin{pmatrix}\boldsymbol{\beta}_1^T & \boldsymbol{\beta}_2^T\end{pmatrix}\begin{pmatrix}0 & 0\\ 0 & \mathbf{x}_2^T\mathbf{x}_2-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$$

$$=\left(n-r-1\right)\sigma^2+\boldsymbol{\beta}_2^T\left[\mathbf{x}_2^T\mathbf{x}_2-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\right]\boldsymbol{\beta}_2$$

ونحصل على القيمة المتوقعة لمتوسط مجموع مربعات البواقي:

$$\begin{aligned}E\left(\frac{RSS_1}{n-r-1}\right)&=\sigma^2+\frac{1}{\left(n-r-1\right)}\boldsymbol{\beta}_2^T\left[\mathbf{x}_2^T\mathbf{x}_2-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\right]\boldsymbol{\beta}_2\\ &=\sigma^2+\frac{\boldsymbol{\beta}_2^T\mathbf{x}_2^T M_1\mathbf{x}_2\boldsymbol{\beta}_2}{\left(n-r-1\right)}\end{aligned}\qquad(7\text{-}3)$$

ونستنتج من هذه النتيجة أن القيمة المتوقعة لمتوسط مجموع مربعات البواقي مقدر متحيز للتباين σ^2 في حالة إسقاط متغيرات مستقلة ذات علاقة، وحجم التحيز هو $\frac{1}{(n-r-1)}\boldsymbol{\beta}_2^T\mathbf{x}_2^T M_1\mathbf{x}_2\boldsymbol{\beta}_2$.

٧-١-٢ إدخال متغيرات مستقلة ليست ذات العلاقة (Inclusion of Irrelevant Variables):

تقدير معالم نموذج الانحدار الخطي في ظل إضافة متغيرات مستقلة غير ذات علاقة:

في هذه الحالة يتم تضمين متغيرات مستقلة ليست ذات علاقة في نموذج الانحدار الخطي. وبافتراض إن النموذج الصحيح الذي يضم المتغيرات المستقلة ذات العلاقة فقط وهو:

$$\mathbf{y}=\mathbf{x}_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon}$$

وتم بناء النموذج التالي:

$$\mathbf{y}=\mathbf{x}_1\boldsymbol{\beta}_1+\mathbf{x}_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$$

حيث إن $\mathbf{x}_1$ مصفوفة بيانات المتغيرات ذات العلاقة من رتبة $n\times(r+1)$ و $\mathbf{x}_2$ مصفوفة بيانات المتغيرات غير ذات العلاقة من رتبة $n\times s$.

إذا ما تم بناء نموذج انحدار خطي يضم (r+1) متغيرات مستقلة ذات علاقة و s $(s \geq 1)$ متغيرات مستقلة غير ذات علاقة نحصل على ما يلي:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y}$$

وبإيجاد التوقع، نحصل على:

$$E\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T E\left(\mathbf{y}\right)$$

وبما أن القيمة المتوقعة لـ y هي $\mathbf{x}_1\boldsymbol{\beta}_1$ فإن

$$\begin{aligned} E\left(\hat{\boldsymbol{\beta}}\right) &= \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{x}_1\boldsymbol{\beta}_1 \\ &= \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\left(\mathbf{x}_1 \quad \mathbf{x}_2\right)\begin{pmatrix}\boldsymbol{\beta}_1 \\ \mathbf{0}\end{pmatrix} \qquad (7\text{-}4) \\ &= \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{x}\begin{pmatrix}\boldsymbol{\beta}_1 \\ \mathbf{0}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_1 \\ \mathbf{0}\end{pmatrix} \end{aligned}$$

ونستنتج من هذه النتيجة أن مقدرات معالم النموذج $\mathbf{x}_1$ مقدرات غير متحيزة لمعالم النموذج الحقيقية $\boldsymbol{\beta}_1$ وأن القيم المتوقعة لمقدرات معالم المتغيرات المستقلة غير ذات العلاقة مساوية للصفر.

تقدير التباين في ظل إضافة متغيرات مستقلة غير ذات علاقة:

مجموع مربعات بواقي النموذج الذي يشتمل على جميع المتغيرات ذات العلاقة وغير ذات العلاقة هو:

$$RSS = \mathbf{y}^T\mathbf{M}\mathbf{y}$$

حيث إن $\mathbf{M} = \mathbf{I} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T$

وللحصول على مقدر التباين، نحسب القيمة المتوقعة لمجموع مربعات البواقي كما يلي:

$$\begin{aligned} E(RSS) &= E\left(\mathbf{y}^T\mathbf{M}\mathbf{y}\right) \\ E(RSS) &= E\left\{\left(\mathbf{x}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}\right)^T\mathbf{M}\left(\mathbf{x}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}\right)\right\} \\ E(RSS) &= E\left\{\boldsymbol{\varepsilon}^T\mathbf{M}\boldsymbol{\varepsilon} + 2\boldsymbol{\beta}_1^T\mathbf{x}_1^T\mathbf{M}\boldsymbol{\varepsilon} + \boldsymbol{\beta}_1^T\mathbf{x}_1^T\mathbf{M}\mathbf{x}_1\boldsymbol{\beta}_1\right\} \end{aligned}$$

ويلاحظ أن الحد $\mathbf{M}\mathbf{x}_1 = \mathbf{0}$ وذلك لما يلي(Ismail,2012):

$$\mathbf{M}\mathbf{x}_1 = \left(\mathbf{I} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\right)\mathbf{x}_1$$

$$\mathbf{M}\mathbf{x}_1 = \mathbf{x}_1 - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{x}_1$$

$$\mathbf{Mx}_1 = \mathbf{x}_1 - \mathbf{x}\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\left(\mathbf{x}_1 \quad \mathbf{x}_2\right)\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix}$$

$$\mathbf{Mx}_1 = \mathbf{x}_1 - \mathbf{x}\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{x}\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix}$$

$$\mathbf{Mx}_1 = \mathbf{x}_1 - \mathbf{x}\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix} = \mathbf{x}_1 - \left(\mathbf{x}_1 \quad \mathbf{x}_2\right)\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix} = \mathbf{x}_1 - \mathbf{x}_1 = \mathbf{0}$$

ومن ثم فإن القيمة المتوقعة لمجموع مربعات البواقي هي:

$$\mathrm{E}\left(\mathrm{RSS}\right) = \mathrm{E}\left(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{M}\boldsymbol{\varepsilon}\right) = \sigma^2 trace(\mathbf{M}) = \sigma^2\left(\mathrm{n}-\mathrm{r}-\mathrm{s}\right)$$

وبذلك يكون مقدر التباين $\mathbf{S}^2$:

$$\mathbf{S}^2 = \frac{\mathrm{RSS}}{\mathrm{n}-\mathrm{r}-\mathrm{s}} \qquad (7\text{-}5)$$

وهو مقدر غير متحيز للتباين σ^2

ويلاحظ من النتائج السابقة أن إسقاط متغيرات ذات علاقة يؤثر في مقدرات معاملات الانحدار والإحصاءات المرتبطة بها، في حين نحصل على مقدرات غير متحيزة في حالة إضافة متغيرات غير ذات علاقة في النموذج.

٧-١-٣ بعض طرق كشف أخطاء توصيف نموذج الانحدار الخطي:

تعد أخطاء توصيف نموذج الانحدار الخطي من المشكلات التي يصعب كشفها نظراً إلى أنه في الواقع العملي لا يُعلم صحة النموذج الذي تم بناؤه. على الرغم من ذلك توجد بعض الاعتبارات يمكن الأخذ بها في الكشف عن توصيف النموذج، منها ما يلي:

١. **مقارنة نتائج النموذج الموفق بالنظريات ذات الصلة:** أفضل طريقة لكشف إسقاط متغيرات ذات علاقة هو مقارنة نتائج النموذج بالنظرية التي على خلفيتها تم بناء النموذج للتأكد من أنه لم يتم إسقاط أي متغير مفسر أو إدخال متغيرات ليست ذات صلة. ففي بعض المجالات، تحدد النظرية المتغيرات المستقلة ذات العلاقة كما هو الحال في بحوث تصميم التجارب. غير أنه في البحوث المشاهدة نادراً ما تكون المتغيرات المفسرة معلومة وهي الحالة التي تحدث فيها أخطاء التوصيف التي تشمل إما إسقاط متغيرات ذات علاقة أو إضافة متغيرات ليست ذات علاقة.

٢. **كشف المتغيرات المستقلة ليست ذات الصلة:** في حال عدم وجود نظريات تم على أساسها بناء النموذج، يمكن من نتائج نموذج الانحدار تحديد المتغيرات المستقلة غير الدالة إحصائياً في مساهمتها في تفسير التباين في المتغير التابع. وفي الدراسات والبحوث المشاهدة الحالة التي يتم قياس عدد كبير من المتغيرات المستقلة المرشحة التي تفسر التباين في المتغير التابع، تستخدم طرق اختيار أفضل نموذج، الموضوعات التي تم استعراضها في الفصل

السادس. تستبعد هذه الطرق بصورة تلقائية المتغيرات المستقلة التي لا تسهم في تفسير المتغير التابع بمستوى معنوي. إلا أنه يجب الانتباه إلى أن هذه الطرق تفترض استيفاء جميع اشتراطات نموذج الانحدار والتي من أهمها صحة شكل العلاقة الدالية بين المتغيرات المستقلة والمتغير التابع وعدم وجود مشاهدات شاذة ومؤثرة في البيانات.

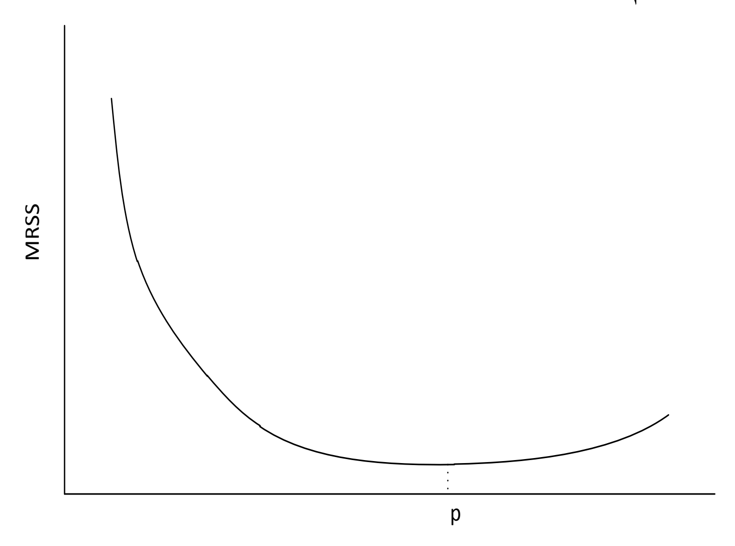
٣. **معامل التحديد المعدل** $\bar{R}^2$ **:** معامل التحديد المعدل هو أحد المؤشرات التي يُعتمد عليها في تقييم مساهمة المتغيرات المستقلة غير ذات العلاقة في تفسير التباين في المتغير التابع. إذ إن إدخال متغير مستقل غير ذي علاقة يسهم في خفض معامل التحديد المعدل. إلا أن هذا المعيار غير كاف في حال كبر حجم العينة (عدد المشاهدات)، إذ إن قيمة معامل التحديد المعدل تقترب من قيمة معامل التحديد غير المعدل كلما ازداد حجم العينة.

٤. **فحص معاملات نموذج الانحدار:** يُعد فحص معاملات النموذج من أهم طرق الكشف عن أخطاء توصيف النموذج. فإشارة معامل المتغير المستقل غير المتوقعة ربما تشير إلى وجود متغيرات ليست ذات علاقة مضمنة في النموذج الموفق أو تشير إلى وجود مشكلة الارتباط الخطي المتعدد المشكلة التي سنتناولها في الجزء التالي.

٥. متوسط مربعات البواقي MRSS: يمثل متوسط مربعات البواقي أحد معايير اختيار أفضل نموذج (انظر الفصل السادس). ويأخذ متوسط مربعات البواقي الصيغة التالية:

$$MRSS = \frac{RSS}{n-p-1}$$

ويتميز متوسط مربعات البواقي بأن قيمته تنخفض كلما زاد عدد المتغيرات المستقلة ذات الصلة وتصل إلى أقل قيمة له عندما يضم النموذج جميع المتغيرات ذات الصلة، ومن ثم تبدأ قيمته في الزيادة بإدخال متغيرات ليست ذات صلة في النموذج (انظر الشكل رقم ٧-١).



شكل (٧-١): العلاقة بين قيمة متوسط مربعات البواقي وعدد المتغيرات المستقلة ذات الصلة

٧-٢ الارتباط الخطي المتعدد (Multicollinearity):

٧-٢-١ مقدمة:

من الاشتراطات الأساسية التي تقوم عليها طريقة المربعات الصغرى عدم وجود علاقة خطية تربط بين أحد المتغيرات المستقلة ومتغير مستقل آخر أو بين أحد المتغيرات المستقلة وأي تركيب خطي بين المتغيرات المستقلة الأخرى. ويشير الارتباط الخطي المتعدد (Multicollinearity) إلى وجود علاقة خطية بين أحد المتغيرات المستقلة ومتغير أو أكثر من المتغيرات المستقلة الأخرى. والارتباط الخطي المتعدد نوعان هما: الارتباط الخطي المتعدد التام (Perfect Multicollinearity) والارتباط الخطي المتعدد المرتفع غير التام (High Imperfect Multicollinearity) .

٧-٢-١-١ الارتباط الخطي التام:

يشير الارتباط الخطي التام إلى أن أحد المتغيرات المستقلة في معادلة الانحدار له ارتباط خطي تام بواحد أو أكثر من المتغيرات المستقلة الأخرى. ورياضياً يقال إن هناك علاقة خطية تامة بين المتغيرات المستقلة $\left(x_1, x_2, \ldots, x_p\right)$ إذا تحقق الشرط التالي:

$$c_0 + c_1 x_{1i} + c_2 x_{2i} + \ldots + c_p x_{pi} = 0 \qquad (7\text{-}6)$$

حيث إن $c_0, c_1, c_2, \ldots, c_p$ قيم ثابتة لا تتساوى جميعها بالصفر.

ومن المعادلة (7.6) يمكن اشتقاق أي متغير مستقل $\left(x_j\ ,\ j=1,2,..,p\right)$ كتركيب خطي لبقية المتغيرات المستقلة على النحو التالي:

المتغير المستقل x_1: $x_{1i} = -\frac{c_0}{c_1} - \frac{c_2 x_{2i}}{c_1} - \ldots - \frac{c_p x_{pi}}{c_1}$ ، حيث أن $c_1 \neq 0$ ،

المتغير المستقل x_2: $x_{2i} = -\frac{c_0}{c_2} - \frac{c_1 x_{1i}}{c_2} - \ldots - \frac{c_p x_{pi}}{c_2}$ ، حيث أن $c_2 \neq 0$ ،

وبالاستمرار نحصل على المتغير x_p كتركيب خطي لبقية المتغيرات:

المتغير المستقل x_p: $x_{pi} = -\frac{c_0}{c_p} - \frac{c_1 x_{1i}}{c_p} - \ldots - \frac{c_{p-1} x_{(p-1),i}}{c_2}$ ، حيث أن $c_p \neq 0$.

وبسبب وجود الارتباط الخطي التام تكون المصفوفة $\left(\mathbf{X}^T\mathbf{X}\right)$ مصفوفة مفردة (Singular)، أي أن محددها يساوى الصفر، ذلك للآتي:

بما أن:

$$(\mathbf{X}^T\mathbf{X}) = \begin{pmatrix} n & \sum x_{1i} & \sum x_{2i} & \dots & \sum x_{pi} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \dots & \sum x_{1i}x_{pi} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 & \dots & \sum x_{2i}x_{pi} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ \sum x_{pi} & \sum x_{pi}x_{1i} & \sum x_{pi}x_{2i} & \dots & \sum x_{pi}^2 \end{pmatrix}$$

فإنه يمكن إيجاد قيم الصف الأخير من هذه المصفوفة مثلاً بدلالة الصفوف الأخرى في صورة علاقة خطية باستخدام المعادلة التالية:

$$x_{pi} = -\frac{c_0}{c_p} - \frac{c_1 x_{1i}}{c_p} - \dots - \frac{c_{p-1} x_{(p-1),i}}{c_2}$$

وبجمع طرفي هذه المعادلة نحصل على:

$$\sum x_{pi} = -\frac{nc_0}{c_p} - \frac{c_1}{c_p}\sum x_{1i} - \dots - \frac{c_{p-1}}{c_p}\sum x_{(p-1),i}$$

وكذلك يمكن الحصول على العنصر $\sum x_{pi}x_{1i}$ بضرب طرفي المعادلة السابقة بـ x_{1i} ثم جمعها لنحصل على:

$$\sum x_{pi}x_{1i} = -\frac{c_0}{c_p}\sum x_{1i} - \frac{c_1}{c_p}\sum x^2_{1i} - \dots - \frac{c_{p-1}}{c_p}\sum x_{(p-1)i}x_{1i}$$

وهكذا يمكن الحصول على كل قيم الصف الأخير من المصفوفة $(\mathbf{X}^T\mathbf{X})$ بدلالة الصفوف الأخرى، وبذلك تكون رتبة هذه المصفوفة أقل من (P+1) - عدد أعمدة المصفوفة- وتكون محددتها مساوية للصفر ولا يمكن إيجاد معكوسها $(\mathbf{X}^T\mathbf{X})^{-1}$ وبالتالي يتعذر استخدام طريقة المربعات الصغرى لتقدير معالم النموذج والإحصاءات الأخرى.

مثال:

البيانات الافتراضية التالية توضح حالة وجود ارتباط خطي تام بين المتغيرين X_1 و X_2:

6	2	1	3	5	4	**X_1**
3	1	0.5	1.5	2.5	2	X_2

ويلاحظ من هذه البيانات أن هناك علاقة خطية تامة بين هذين المتغيرين ذلك لأن: $x_{1i} = 2\,x_{2i}$

ويمكن حساب المصفوفة $(\mathbf{X}^T\mathbf{X})$ كما يلي:

$$(\mathbf{X}^T\mathbf{X}) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 5 & 3 & 1 & 2 & 6 \\ 2 & 2.5 & 1.5 & 0.5 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 4 & 2 \\ 1 & 5 & 2.5 \\ 1 & 3 & 1.5 \\ 1 & 1 & 0.5 \\ 1 & 2 & 1 \\ 1 & 6 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 21 & 10.5 \\ 21 & 91 & 45.5 \\ 10.5 & 45.5 & 22.75 \end{pmatrix}$$

ويتضح من بيانات المصفوفة $(\mathbf{X}^T\mathbf{X})$ العلاقة الخطية التامة بين المتغيرين x_1 وx_2 حيث يمكن إيجاد قيم عناصر الصف الأخير بدلالة الصف الثاني كما يلي:

نعلم أن:

$$x_{1i} - 2\,x_{2i} = 0$$

وبجمع طرفي هذه المعادلة نحصل على:

$$\sum x_{2i} = \tfrac{1}{2}\sum x_{1i}$$

وبضرب طرفي المعادلة بـ x_{1i} وثم جمعهما نحصل على:

$$\sum x_{1i}^2 = 2\sum x_{2i}x_{1i}$$

وهكذا نجد أن قيمة أية عنصر من عناصر الصف الأخير ما هو إلا عبارة عن نصف قيمة العنصر المقابل من الصف الثاني، وكذلك نجد قيمة أية عنصر من العمود الثالث والأخير مساو لنصف قيمة العنصر المناظر له من العمود الثاني. وبذلك نجد أن رتبة المصفوفة $(\mathbf{X}^T\mathbf{X})$ تساوي (٢) أقل من عدد الأعمدة (٣ أعمدة) وتبلغ محددتها صفراً، أي: $|\mathbf{X}^T\mathbf{X}| = 0$. ومن ثم لا يمكن إيجاد معكوس هذه المصفوفة اللازم لحساب مقدرات المربعات الصغرى.

ومن أسباب بروز مشكلة الارتباط الخطي المتعدد التام نذكر ما يلي:

- التعريف الخاطئ للمتغيرات الصورية (Dummy variables) كإدراج k متغير صوري بعدد فئات المتغير التصنيفي في نموذج الانحدار بدلاً من تعريف (k-1) متغير صوري؛ المشكلة التي تعرف بمصيدة المتغيرات الصورية (Dummy variables trap).
- إدراج متغير مستقل ذي قيم ثابتة.

- إدراج متغير مستقل واحد مرتين باستخدام وحدتي قياس مختلفة، كإدراج الوزن بالرطل وبالكيلوجرام مرة أخرى.

وفي الواقع العملي نجد أن مشكلة الارتباط الخطي التام نادر الحدوث، ولحسن الحظ نجد أن معظم برامج الإحصاء الجاهزة تكشف عن وجود هذه المشكلة بمجرد إعطاء الأمر المحدد للحاسب الآلي لحل النموذج المراد تقديره.

٧-٢-١-٢ الارتباط الخطي المتعدد المرتفع:

يشير الارتباط الخطي المتعدد المرتفع إلى الحالة التي يكون فيها بين اثنين أو أكثر من المتغيرات المستقلة ارتباطا قوياً ولكنه ليس تاماً مما يجعل من الصعب عزل تأثيراتها الفردية على المتغير التابع. وهذا النوع من الارتباط الخطي المتعدد هو الذي يهم الباحث المستخدم لأسلوب تحليل الانحدار الخطي. وتجدر الإشارة إلى أن مشكلة الارتباط الخطي المتعدد تحدث بشكل متكرر في البيانات المشاهدة في حقول الدراسات الإنسانية والاجتماعية. ويرجع أسباب ظهور مشكلة الارتباط الخطي المتعدد إلى الآتي:

- ميل بعض المتغيرات للتغير سوياً. فمثلاً نجد أن متغيرات دخل الموظف وسنوات خبرته وعمره ومرتبته الوظيفية غالباً ما تتغير سوياً ويوجد بينها ارتباط موجب قوي.
- قلة عدد المشاهدات مقارنة بعدد المتغيرات المضمنة في النموذج.
- إدراج متغيرات متباطئة (Lagged variables) كمتغيرات مفسرة، كإدراج سعر محصول ما لنفس الفترة وسعره في فترة سابقة كمتغيرين مفسرين لكمية إنتاج المحصول. حيث يأخذ النموذج في هذه الحالة الصيغة التالية: $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t, \quad t = 1,2,...,n$، حيث أن y_t كمية إنتاج المحصول، x_t سعر المحصول في الفترة t و x_{t-1} سعر المحصول في الفترة السابقة.

٧-٢-٢ النتائج المترتبة على وجود الارتباط الخطي في نموذج الانحدار الخطي:

إذا كان الارتباط الخطي مرتفع ولكنه غير تام يمكننا حساب مقدرات المربعات الصغرى ذلك لأن المصفوفة $(\mathbf{X}^T\mathbf{X})$ في هذه الحالة غير مفردة وأن محددتها تختلف عن الصفر ولكنها قريبة منه. ويترتب على وجود الارتباط الخطي المتعدد الآتي:

- تظل مقدرات طريقة المربعات الصغرى لها "أقل تباين" من بين مجموعة المقدرات الأخرى غير المتحيزة مع ملاحظة أن "أقل تباين" لا يعني "تباين قليل".
- بما أن قيمة محددة المصفوفة $|\mathbf{X}^T\mathbf{X}|$ في ظل الارتباط الخطي المتعدد تكون قريبة من الصفر، فإن قيم العناصر القطرية لمعكوس المصفوفة - $(\mathbf{X}^T\mathbf{X})$ - $(\mathbf{X}^T\mathbf{X})^{-1}$ - ستكون كبيرة ومن ثم تكون قيم الأخطاء المعيارية لمعاملات نموذج الانحدار المقدرة كبيرة أيضاً. ولتوضيح تأثير الارتباط الخطي المتعدد في زيادة قيم التباين والتغاير ومن

ثم في زيادة قيم الأخطاء المعيارية لمعاملات نموذج الانحدار المقدرة، نقوم بحساب النموذج المعياري بتحويل كل من المتغير التابع والمتغيرات المستقلة متغيرات معيارية لنحصل على النموذج التالي:

$$\widehat{\boldsymbol{\beta}}_z = \left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{y}^o \qquad (7\text{-}7)$$

حيث أن:

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1p} \\ W_{21} & W_{22} & \dots & W_{2p} \\ \dots & \dots & \dots & \dots \\ W_{n1} & W_{n2} & \dots & W_{np} \end{pmatrix}, \quad \mathbf{y}^0 = \begin{pmatrix} y_1^o \\ y_2^o \\ \dots \\ y_n^o \end{pmatrix}$$

$$W_{ij} = \frac{x_{ij} - \overline{x}_j}{S_{jj}^{\frac{1}{2}}}, \qquad i=1,2,...,n,\ \ j=1,2,...,p$$

$$y_i^o = \frac{y_i - \overline{y}}{S_T^{\frac{1}{2}}}, \quad i=1,2,...,n$$

$$S_T = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \ \text{و}\ S_{jj} = \sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2$$

وبما أن المتغير التابع والمتغيرات المستقلة متغيرات معيارية فإن المصفوفة $\left(\mathbf{W}^T\mathbf{W}\right)$ هي مصفوفة معاملات الارتباط بين المتغيرات المستقلة، والمتجه $\mathbf{W}^T\mathbf{y}^0$ هو متجه معاملات ارتباط المتغير التابع مع كل من المتغيرات المستقلة ، أي:

$$\mathbf{W}^T\mathbf{W} = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{21} & 1 & r_{23} & \dots & r_{2p} \\ r_{31} & r_{32} & 1 & \dots & r_{3p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & r_{p3} & \dots & 1 \end{pmatrix}, \quad \mathbf{W}^T\mathbf{y}^0 = \begin{pmatrix} r_{1y} \\ r_{2y} \\ r_{3y} \\ \dots \\ r_{py} \end{pmatrix}$$

وتسمى العناصر القطرية لمعكوس المصفوفة $\mathbf{W}^T\mathbf{W}$ ، $\left(\mathbf{W}^T\mathbf{W}\right)^{-1}$ ، بعوامل تضخم التباين C_{jj} ، أي:

$$\text{VIF}_j = C_{jj} = \frac{1}{1 - R_j^2}, \quad j=1,2,...,p \qquad (7\text{-}8)$$

حيث إن R_j^2 معامل التحديد لنموذج انحدار x_j على بقية المتغيرات المستقلة $(p-1)$.

ونحصل على تباين معاملات الانحدار كما يلي:

$$\operatorname{var}\left(\widehat{\beta}_{zj}\right)=\sigma_z^2\left(\mathbf{W}^T\mathbf{W}\right)_{jj}^{-1}=\sigma_z^2 C_{jj}=\sigma_z^2\left(1-R_j^2\right)^{-1} \quad (7\text{-}9)$$

ويتضح من هذه المعادلة أنه كلما كانت قيمة R_j^2 قريبة من الواحد كلما كانت قيمة التباين كبيرة. ويترتب على كبر الأخطاء المعيارية أن تكون قيم إحصاء (t) صغيرة مما يقود إلى عدم معنوية هذه المقدرات واتساع فترة الثقة لمعالم النموذج.

- على الرغم من أن كبر حجم الأخطاء المعيارية لمعلمات النموذج المقدرة هو الأثر المباشر الناتج عن وجود الارتباط الخطي المتعدد، إلا أنه ليس في كل حالة نجد فيها قيم أخطاء معيارية كبيرة يرجع سببها لوجود هذه المشكلة. حيث توجد أسباب أخرى تسهم في تضخم الأخطاء المعيارية كصغر حجم العينة أو إدراج متغيرات مفسرة ذات تباين قليل.
- عدم دقة واستقرار المعلمات المقدرة من عينة إلى أخرى ويلاحظ أن أي تغيير طفيف في العينة كحذف أو إضافة مشاهدات أو مشاهدات قليلة يؤدي إلى تغيير كبير في حجم وإشارات معاملات النموذج المقدر.

٧-٢-٣ طرق الكشف عن وجود الارتباط الخطي المتعدد:

توجد مقاييس عديدة تستخدم للكشف عن وجود الارتباط الخطي المتعدد بين المتغيرات المستقلة. فيما يلي نستعرض بعضاً منها:

١. من أهم المؤشرات التي تدل على وجود مشكلة الارتباط الخطي المتعدد هو أن يكون معظم أو كل معاملات الانحدار الجزئية غير معنوية إحصائياً و/أو ذات إشارات مخالفة للمتوقعة أو المفترضة على الرغم من كبر حجم معامل التحديد (R^2) ومعنوية الانحدار ككل.

٢. عندما يوجد ارتباط خطي متعدد فإن أي تغيير طفيف في العينة كحذف أو إضافة مشاهدة أو مشاهدات قليلة يؤدي إلى تغيير كبير في حجم وإشارات معاملات النموذج.

٣. فحص مصفوفة معاملات الارتباط الخطي البسيط بين أزواج المتغيرات المستقلة $\mathbf{R}_{xx}$ حيث

$$\mathbf{R}_{xx} = \begin{pmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_p} \\ & 1 & r_{x_2x_3} & \dots & r_{x_2x_p} \\ & & 1 & \dots & r_{x_3x_p} \\ \dots & \dots & \dots & \dots & \dots \\ & & & \dots & r_{x_{p-1},x_p} \\ & & & \dots & 1 \end{pmatrix} \qquad (7\text{-}10)$$

حيث إن:

$$r_{x_hx_j} = \frac{\sum(x_{hi}-\bar{x}_h)(x_{ji}-\bar{x}_j)}{\sqrt{\sum(x_{hi}-\bar{x}_h)^2}\sqrt{\sum(x_{ji}-\bar{x}_j)^2}}, \qquad h,j=1,2,\dots,p$$

وبملاحظة قيم معاملات الارتباط، إذا وجد ارتباط قوي بين أي متغيرين مستقلين دلّ ذلك على احتمال وجود ارتباط خطي متعدد. ولكن يجب ملاحظة أن ضعف العلاقة الزوجية بين المتغيرات المستقلة لا يعني غياب المشكلة؛ إذ يمكن أن يكون هناك علاقة خطية أو تركيب خطي بين أحد المتغيرات المستقلة ومتغيرين أو أكثر من بقية المتغيرات المستقلة.

٤. من الاختبارات المناسبة التي تستخدم للكشف عن وجود مشكلة الارتباط الخطي المتعدد هو بناء عدد (p) نموذج انحدار لكل متغير مستقل على بقية المتغيرات المستقلة كما يلي:

$$x_1 = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p + e_i$$
$$x_2 = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \dots + \hat{\beta}_p x_p + e_i$$
$$\dots$$
$$x_p = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_{p-1} x_{p-1} + e_i$$

فإذا كانت قيمة أحد معاملات التحديد (R^2) لهذه النماذج تقترب من الواحد الصحيح دلّ ذلك على وجود ارتباط خطي متعدد. ويعد هذا الاختبار أفضل من اختبار فحص مصفوفة معاملات الارتباط الخطي البسيط بين أزواج المتغيرات المستقلة.

٥. عامل تضخم التباين (Variance Inflation Factor {VIF}): من الطرق الأساسية الواسعة الاستخدام للكشف عن وجود الارتباط الخطي المتعدد هو عامل تضخم التباين (VIF). وتقيس عوامل التضخم مدى تضخم تباينات معاملات الانحدار المقدرة في وجود الارتباط الخطي. فإذا كانت قيمة عامل تضخم التباين (VIF_j) أكبر من (١٠) كان ذلك دلالة على وجود ارتباط خطي متعدد مرتفع (Bowerman et al, 2005, p.224;Kim and Timm, 2007; Neter et al,1990, p409). ويكافئ هذا الحد عندما تكون قيمة معامل التحديد لنموذج انحدار المتغير المستقل رقم (j) على بقية المتغيرات المستقلة مساوية لـ ٠,٩٠ ($R^2_j = 0.90$).

ويلاحظ الآتي على صيغة عامل تضخم التباين:

- يأخذ عامل تضخم التباين قيماً غير سالبة، أي: $VIF_j \geq 0$
- في حالة وجود ارتباط خطي تام بين المتغير المستقل X_j وبقية المتغيرات المستقلة $(R^2_j=1)$ فإن عامل تضخم التباين يتخذ قيمةً لا نهائية، لأن: $VIF_j = \frac{1}{1-1} = \frac{1}{0} = \infty$

وفي حالة عدم وجود ارتباط خطي بين المتغير المستقل X_j وبقية المتغيرات المستقلة $(R_j^2 = 0)$ فإن قيمة عامل تضخم التباين تكون مساويةً للواحد الصحيح.

- تستخدم عوامل تضخم التباين (VIF's) لقياس مدى بعد مقدرات المربعات الصغرى عن قيمها الحقيقية. حيث تأخذ القيم المتوقعة لمجموع مربعات الفروق بين معاملات الانحدار المقدرة وقيمها الحقيقية الصيغة التالية:

$$E\{\sum_{j=1}^{p}(\hat{\beta}_j - \beta_j)^2\} = \sigma^2 \sum_{j=1}^{p} VIF_j$$

وكما أشرنا في النقطة السابقة في حالة عدم وجود ارتباط خطي بين المتغيرات المستقلة، تكون قيم عوامل تضخم التباين جمعيها مساويةً للواحد الصحيح. وفي هذه الحالة تأخذ المعادلة أعلاه الصيغة التالية:

$$E\{\sum_{j=1}^{p}(\hat{\beta}_j - \beta_j)^2\} = \sigma^2 \sum_{j=1}^{p} VIF_j = \sigma^2 p$$

ومن ثم يمكن حساب النسبة التالية:

$$\frac{\sigma^2 \sum_{j=1}^{p} VIF_k}{p\sigma^2} = \frac{\sum_{j=1}^{p} VIF_j}{p} \qquad (7\text{-}11)$$

ويلاحظ أن هذه النسبة عبارة عن متوسط قيم عوامل تضخم التباين لمعاملات الانحدار. فإذا كانت المتغيرات المستقلة متعامدة أي لا يوجد بينها ارتباط خطي فإن هذه النسبة تساوى واحد صحيح. ولذلك نجد أنه كلما زادت قيمة متوسط عوامل تباين التضخم عن الواحد الصحيح كلما دل ذلك على وجود الارتباط الخطي بين المتغيرات المستقلة.

- تستخدم بعض حزم برامج الإحصاء الجاهزة معكوس عامل التضخم للكشف عن وجود الارتباط الخطي بين المتغير المستقل X_j والمتغيرات المستقلة الأخرى وتحديد دخول المتغير للنموذج أم لا. ويعرف هذا المقياس بالتحمل (Tolerance) ويتم حسابه حسب الصيغة التالية:

$$\text{Tolerance} = \frac{1}{\text{VIF}_j} = 1 - R_j^2 \qquad (7\text{- }12)$$

وقيم التحمل التي تستخدم بواسطة هذه البرامج كحد أدنى لدخول أي متغير مستقل النمـوذج هـي: ٠,٠١، ٠,٠٠١، أو ٠,٠٠٠١.

- يأخذ تباين مقدر المربعات الصغرى للمعامل ($\hat{\beta}_j$) الصيغة التالية:

$$\text{var}(\hat{\beta}_j) = \frac{S^2 \text{VIF}_j}{(n-1)S_j^2} = \frac{1}{1\text{-}\ R_j^2}\,\frac{S^2}{(n-1)S_j^2} \qquad (7\text{- }13)$$

حيث إن:

S^2= تباين التقدير لنموذج انحدار Y على كل المتغيرات المستقلة.

S_j^2= تباين المتغير المستقل X_j، أي: $S_j^2 = \frac{\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2}{n-1}$

VIF_j = عامل تضخم تباين المتغير المستقل X_j.

وتوضح هذه المعادلة أثر الارتباط الخطي المتعدد على دقة تقدير المعلمة ($\hat{\beta}_j$)، إذ يلاحظ أن قيمـة تبـاين المعامـل $\hat{\beta}_j$ تزداد مع زيادة قيمة R_j^2.

- إن عامل تضخم التباين رقم (j) هو العنصر القطري رقم (j) لمعكوس مصفوفة معـاملات الارتبـاط البسـيط بـين المتغيرات المستقلة $\mathbf{R}^{-1}_{XX}$ (Theil, 1971, p. 166)، أي:

$$\text{VIF}_j = \text{diag}\left(\mathbf{R}^{-1}_{XX}\right)_{jj} \qquad (7\text{-}14)$$

- طوّر فوكس ومونيت (Fox & Monette, 1992, pp178-183) عامل تضخم التباين للكشـف عـن الارتبـاط الخطـي المتعدد في حالة وجود ارتباط متوقع بين المتغيرات المستقلة كاشتمال نموذج الانحـدار عـلى متغـيرات صـورية أو متغيرات قوة في حالة معادلات الانحدار متعـدد الحـدود باسـتخدام مـا أسـماه بعامـل تضـخم التبـاين المعمـم (Generalized Variance Inflation Factor). حيث يمكن إعادة كتابة نموذج الانحدار كما يلي:

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \boldsymbol{\varepsilon}$$

حيث إن:

$\mathbf{x}_1$= مصفوفة البيانات التي تحتوي على k من المتغيرات المرتبطة (مثال: متغيرات صورية).

$\mathbf{x}_2$ = مصفوفة البيانات التي تحتوي على بقية المتغيرات (p-k). واستخدم فوكس ومونيت الصيغة التالية لحساب عامل تضخم التباين المعمم:

$$GVIF_1 = \frac{|\mathbf{R}_{11}| \times |\mathbf{R}_{22}|}{|\mathbf{R}|} \qquad (7\text{-}15)$$

حيث إن: $|\mathbf{R}_{11}|$ = محددة مصفوفة معاملات الارتباط لـ x_1، $|\mathbf{R}_{22}|$= محددة مصفوفة معاملات الارتباط لــ x_2 و $|\mathbf{R}|$= محددة مصفوفة معاملات الارتباط لجميع المتغيرات.

٦. **قيم الجذر الكامنة**[*] (Eigenvalues): يعد حساب قيم الجذر الكامنة/المميزة (Eigenvalues) لمصفوفة الارتباط بين معاملات نموذج الانحدار من المؤشرات المهمة التي تستخدم لقياس الارتباط الخطي المتعدد. وتتلخص هذه الطريقة في الآتي:

- يتم أولاً حساب المصفوفة $\mathbf{z}$، حيث: $\mathbf{Z} = \left(\mathbf{x}^T\mathbf{x}\right)$
- يتم حساب المصفوفة القطرية $\mathbf{S}$ كما يلي: $\mathbf{S} = \text{diag}(\mathbf{x}^T\mathbf{x})^{-0.5}$
- ويتم حساب المصفوفة $\mathbf{SZS}$ التي تحتوي عناصرها القطرية على الواحد الصحيح كما يلي:

$$\mathbf{SZS}=\mathbf{S}(\mathbf{x}^T\mathbf{x})\mathbf{S} \qquad (7\text{-}16)$$

- وباستخدام طرق المصفوفات يتم إيجاد القيم الكامنة/المميزة للمصفوفة $\mathbf{SZS}$. فإذا اتضح أن هناك قيمة كامنة/ مميزة قريبة من الصفر دلّ ذلك على وجود ارتباط خطي شبه تام، أما إذا كانت قيمة أي من القيم الكامنة مساوية للصفر فيعني ذلك وجود ارتباط خطي تام بين المتغيرات المفسرة. كما تستخدم القيم الكامنة لحساب نوعين من الإحصاءات المساعدة في الكشف عن وجود الارتباط الخطي المتعدد هما: مؤشر الحالة (Condition Index) ورقم الحالة (Condition Number). ويتم حساب مؤشر الحالة (CI_j) كما يلي:

$$CI_j = \sqrt{\frac{\lambda_{max}}{\lambda_j}}, \quad j = 1,2,...,p \qquad (7\text{-}17)$$

حيث إن:

λ_{max}= قيمة أكبر قيمة كامنة.

λ_j = القيمة الكامنة رقم j

أما رقم الحالة (Condition Number) فيتم حسابه كما يلي:

---

* للمزيد حول هذا الموضوع يرجى الرجوع إلى Belsley, Kuh, and Welsch , 1980, pp85-191

$$CN = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \qquad (7\text{-}18)$$

وحسب جونستون (Johnston, 1984, p250) إذا كانت قيمة CN ما بين ٢٠ إلى ٣٠ دلّ ذلك وجود ارتباط خطي مرتفع. في حين يقترح بيلسلي وآخرون (Belsley et al ,1980, p105) وكوهين وآخرون (Cohen, Cohen West and Aiken, 2003) إذا كانت قيمة CN ما بين ٣٠ و١٠٠ كان ذلك دلالة على وجود ارتباط خطي متعدد مرتفع جداً.

٧-٢-٤ بعض الحلول المقترحة لعلاج مشكلة الارتباط الخطي المتعدد:

- حيث إن مشكلة الارتباط الخطي المتعدد هي في الأساس مشكلة بيانات عينة فقد يكون أحد الحلول هو الحصول على بيانات إضافية عن طريق زيادة حجم العينة (Gujarati, 1988; Wooldridge, 2009; Brooks, 2002). ويلاحظ هنا أن زيادة عدد المشاهدات (n) تسهم في خفض قيم الأخطاء المعيارية لو لم تتغير القيم الأخرى. على الرغم من صعوبة زيادة حجم العينة خاصة في حالات تصميم التجارب والمسوحات الاجتماعية المرتبطة بظواهر محددة بفترة زمنية، إلا أنه بالإمكان زيادة حجم العينة في بعض الحالات كالاستفادة من البيانات الثانوية.

- استبعاد أحد المتغيرات ذات الارتباط المرتفع مع ملاحظة أن عملية إسقاط المتغيرات قد تؤدي في بعض الحالات إلى عملية تحيز في التقديرات خاصة إذا كان المتغير المستبعد ذا أهمية أساسية في تفسير التغير في المتغير التابع.

- ومن الحلول المقترحة استخدام معلومات قبلية (Priori Information) حول العلاقات بين المتغيرات المفسرة. فمثلاً نجد أن هناك علاقة بين المرتبة الوظيفية ومدة خبرة الموظف في العمل؛ فبدلاً من إدخال متغير المرتبة (X_1) والخبرة (X_2) كمتغيرين مفسرين ضمن متغيرات مفسرة أخرى للأجر الذي يتقاضاه الموظف (Y_i)، يمكن دمج هذين المتغيرين في متغير واحد إذا أمكن الحصول على معلومة تقريبية تفيد بأن قيمة معلمة الخبرة مثلاً تساوى قرابة ربع معلمة المرتبة. وبالحصول على هذه المعلومة يتم بناء النموذج التالي:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

وبما أن $\beta_2 = 0.25\beta_1$ فإن:

$$y_i = \beta_0 + \beta_1 x_{1i} + 0.25\beta_1 x_{2i} + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 (x_{1i} + 0.25 x_{2i}) + \varepsilon_i$$

ويعاب على هذا الحل صعوبة تحديد الأثر الفردي للمتغيرين على المتغير التابع.

- ومن بين الحلول المهمة والعملية تقليل عدد المتغيرات المستقلة ذات الارتباط المرتفع باستخدام تحليل المكونات الأساسية (Principal Component Analysis) أو التحليل العاملي* (Factor Analysis). وتهدف هاتان الطريقتان إلى تحويل المتغيرات المترابطة إلى عدد أقل تسمى بالعوامل (Factors) في حالة التحليل العاملي وبالمكونات الأساسية (Principal Components) في حالة تحليل المكونات الأساسية. بحيث يكون لكل عامل/مكون من هذه العوامل/المكونات دالةً تربطه ببعض أو كل هذه المتغيرات ومن ثم يتم استخدام المتغيرات الجديدة غير المرتبطة بعضها مع بعض كمتغيرات مفسرة جديدة للمتغير التابع.

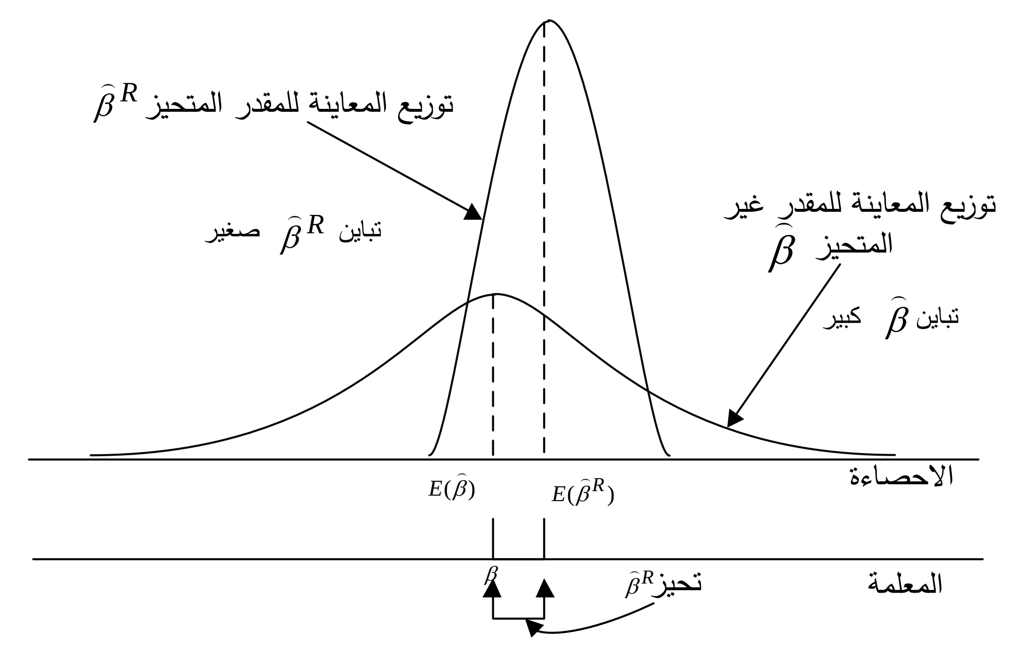
- **انحدار التل (Ridge Regression):** انحدار التل هو أحد الطرق التي اقترحها كل من هيورل وكينارد (Hoerl and Kennard, 1970a, 1970b) لعلاج مشكلة الارتباط الخطي المتعدد. حيث يتم تقدير معالم نموذج الانحدار بإجراء تعديل على طريقة المربعات الصغرى. وباستخدام طريقة انحدار التل يتم الحصول على مقدرات ($\hat{\beta}^R$) متحيزة، إلا أنها أكثر استقراراً من المقدرات غير المتحيزة، وإن احتمال أن يكون قيمة ($\hat{\beta}^R$) قريبة من القيمة الحقيقية للمعلمة أكبر بكثير من احتمال قرب المقدر غير المتحيز ($\hat{\beta}$) لقيمة المعلمة الحقيقية β. ويتضح من الشكل رقم (٧-١) أن المقدر $\hat{\beta}$ غير متحيز غير أنه غير مستقر (imprecise) لكبر حجم تباينه، في حين أن مقدر انحدار التل $\hat{\beta}^R$ أكثر استقراراً لصغر حجم تباينه لكن لديه تحيزاً قليلاً.

ويتم قياس الأثر المشترك للتحيز (bias) والتباين بحساب القيمة المتوقعة لمربع انحراف المقدر المتحيز ($\hat{\beta}^R$) عن القيمة الحقيقية للمعلمة (β) ويسمى هذا المقياس بمتوسط مربعات الخطأ (Mean squared error) ويتم حسابه كما يلي:

$$E(\hat{\beta}^R - \beta)^2 = \sigma^2\{\hat{\beta}^R\} + (E\{\hat{\beta}^R\} - \beta)^2 \qquad (7\text{-}19)$$

وتوضح المعادلة (7.19) أن متوسط مربعات الخطأ يساوى تباين المقدر ($\hat{\beta}^R$) زائداً مربع التحيز مع ملاحظة أن متوسط مربعات الخطأ يساوي تباين المقدر إذا كان غير متحيز.

* للمزيد حول تحليل المكونات الأساسية و التحليل العاملي انظر (Everitt and Dunn, 2010)



شكل رقم (٧-١): توزيع المعاينة للمقدر المتحيز والمقدر غير المتحيز

المصدر: Paulson (2007).

مقدرات انحدار التل:

لإجراء انحدار التل المعياري (Standardized ridge regression) تتبع الخطوات التالية (DeMaris, 2004;Birkes and Dodge, 1993; Paulson, 2007):

- يتم أولاً تحويل المتغير التابع والمتغيرات المستقلة باستخدام طريقة تحويلة الارتباط (Correlation transformation) على النحو التالي:

المتغير التابع: $y_i' = \frac{1}{\sqrt{n-1}} \frac{(y_i - \bar{y})}{S_y}$ ،

حيث إن: $\bar{y} = \frac{\sum y_i}{n}$ ، و $S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$

المتغيرات المستقلة: $r=1,2,...,p$ ، $x_{ri}' = \frac{1}{\sqrt{n-1}} \frac{(x_{ri} - \bar{x}_r)}{S_{x_r}}$ ،

حيث إن: $S_r = \sqrt{\frac{\sum (x_{ri} - \bar{x}_r)^2}{n-1}}$ و $r = 1,2,\ldots,p$ $\quad \bar{x}_r = \frac{\sum x_{ri}}{n}$

وبما أن:

$$\left(\mathbf{x}'^{T}\mathbf{x}'\right) = \mathbf{R}_{xx}$$

و

$$\left(\mathbf{x}'^{T}\mathbf{y}'\right) = \mathbf{R}_{xy}$$

وبإضافة قيمة ثابتة صغيرة - c- للعناصر القطرية لمصفوفة معاملات الارتباط البسيط $\mathbf{R}_{xx}$ تنخفض قيم معاملات الارتباط بمقدار $\frac{1}{(1+c)}$ ومن ثم سيكون ناتج مصفوفة معاملات الارتباطات كالتالي:

$$\mathbf{R}_{xx,c} = \mathbf{R}_{xx}\left(\mathbf{I} + \mathbf{D}_c\right)$$

حيث إن $\mathbf{D}_c$ مصفوفة قطرية قيمة أي عنصر منها مساوية للثابت c ورتبتها $p \times p$ و$\mathbf{I}$ مصفوفة وحدة رتبتها أيضاً $p \times p$ والتي يمكن استخدامها للحصول على مقدرات متحيزة لمعالم نموذج الانحدار. ومن ثم فإنه يمكن الحصول على مقدرات التل المعيارية على النحو التالي:

$$\hat{\boldsymbol{\beta}}^{R} = \left(\mathbf{R}_{xx} + \mathbf{cI}\right)^{-1} \mathbf{R}_{xy} \qquad (7\text{-}20)$$

حيث إن:

$\hat{\boldsymbol{\beta}}^{R} =$ متجه معاملات انحدار التل المعيارية.

$\mathbf{R}_{xx} =$ مصفوفة معاملات الارتباط البسيط بين أزواج المتغيرات المستقلة.

$\mathbf{R}_{xy} =$ متجه معاملات الارتباط البسيط بين المتغير التابع وكل من المتغيرات المستقلة، أي أن:

$$\mathbf{R}_{xy} = \begin{pmatrix} r_{yx1} \\ r_{yx_2} \\ \cdot \\ r_{yx_p} \end{pmatrix}$$

$\mathbf{c}$ = ثابت التحيز وتتراوح قيمته ما بين الصفر والواحد الصحيح.

$\mathbf{I} =$ مصفوفة وحدة من الرتبة $p \times p$.

ولإيجاد قيم معاملات نموذج الانحدار الأصلي تستخدم العلاقة التالية:

$$\hat{\beta}_r = \frac{S_y}{S_{x_r}} \hat{\beta}_r^R, \quad r = 1,2,...,p \qquad (7\text{-}21)$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - ... - \hat{\beta}_p \bar{x}_p$$

عامل تضخم التباين لمعاملات انحدار التل المعيارية: إن عوامل تضخم التباين عبارة عن قيم عناصر المصفوفة التالية:

$$\left(\mathbf{R}_{xx} + c\mathbf{I}\right)^{-1} \mathbf{R}_{xx} \left(\mathbf{R}_{xx} + c\mathbf{I}\right)^{-1} \qquad (7\text{-}22)$$

مجموع مربعات البواقي (Residual Sum of Squares): يتم حساب مجموع مربعات البواقي حسب الصيغ التالية:

$$RSS_R = \sum_{i=1}^{n} (y_i' - \hat{y}_i')^2 \qquad (7\text{-}23)$$

حيث إن:

$$\hat{y}_i' = \hat{\beta}_1^R x_{1i}' + ... + \hat{\beta}_p^R x_{pi}'$$

معامل تحديد انحدار التل: يتم حساب معامل التحديد حسب الصيغة التالية:

$$R_R^2 = 1 - RSS_R \qquad (7\text{-}24)$$

ملاحظات على طريقة انحدار التل:

١. تعكس قيمة الثابت (c) مقدار التحيز في المقدرات ويلاحظ أنه عندما تكون قيمة الثابت مساوية للصفر نحصل على مقدرات المربعات الصغرى الاعتيادية. وعندما تكون قيمة الثابت أكبر من الصفر نحصل على مقدرات متحيزة إلا أنها أكثر استقراراً من مقدرات المربعات الصغرى الاعتيادية.

٢. يعاب على طريقة انحدار التل صعوبة تحديد قيمة (c) المثلى. ولتحديد قيمة التحيز c التي تعطي أفضل نموذج، يستخدم عادة الرسم البياني لقيم معاملات انحدار التل (المحور الصادي) مع قيم مختلفة لثابت التحيز ذات مسافات متساوية (المحور الأفقي). ويعرف الشكل الناتج بــ Ridge trace. وكما يؤخذ في الاعتبار قيمة عامل تضخم التباين للتأكد من حل مشكلة الارتباط الخطي. فإذا أظهر الشكل استقراراً في قيم معاملات الانحدار وانخفاض قيم عوامل تضخم التباين عند قيم محددة لثابت التحيز، يتم اختيار أحد النماذج المناظرة بصورة تحكمية. وتوجد طرق أخرى تحليلية لاختيار قيمة التحيز من أهمها المعادلة التالية التي طورها كل من هورل وكينارد وبالدون (Hoerl, Kennard, and Baldwin, 1975) :

$$c = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^{R^T} \hat{\boldsymbol{\beta}}^R} \qquad (7\text{-}25)$$

حيث إن: p عدد المتغيرات المستقلة، و $\hat{\sigma}^2$ مقدر التباين، و $\hat{\boldsymbol{\beta}}$ مقدرات معالم النموذج المعيارية. كما طور هورل وكينارد (Hoerl and Kennard, 1976) طريقة تتابعية للوصول إلى القيمة المثلى لثابت التحيز. وتتلخص الطريقة في الخطوات التالية:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &: \ c_0 = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^{R^T}\hat{\boldsymbol{\beta}}^R} \\
\hat{\boldsymbol{\beta}}^R_{c_0} &: \ c_1 = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^{R^T}_{c_0}\hat{\boldsymbol{\beta}}^R_{c_0}} \\
\hat{\boldsymbol{\beta}}^R_{c_1} &: \ c_2 = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^{R^T}_{c_1}\hat{\boldsymbol{\beta}}^R_{c_1}}
\end{aligned} \tag{7-26}$$

ويتم الاستمرار في بناء النموذج إلى أن يكون التغيير في قيمة التحيز أكبر من القيمة المحددة في المعادلة التالية:

$$\frac{c_{j+1} - c_j}{c_j} > 20\left(\frac{\text{trace}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}}{p}\right)^{-1.3} \tag{7-27}$$

٧-٢-٥ بعض حالات الارتباط الخطي المتعدد التي يمكن تفاديها:

هناك بعض حالات الارتباط الخطي التي يمكن تفاديها نذكر منها ما يلي:

- تنشأ مشكلة الارتباط الخطي أحياناً في النماذج التي تحتوى على متغيرات قوة كنماذج انحدار الدرجة الثانية فما فوق (Polynomial regression). وينصح في هذه الحالة بإجراء توسيط للمتغير (Centering) أو المتغيرات وذلك بطرح قيمة الوسط الحسابي للمتغير من قيمة أي مشاهدة ومن ثم تتم عملية رفع القوة للدرجة أو الدرجات المطلوبة وإجراء نموذج الانحدار كالمعتاد.

- تنشأ مشكلة الارتباط الخطي أحياناً من وجود مشاهدات شاذة في المتغيرات المستقلة أو في المتغير التابع، ولذلك لا بد من فحص البيانات للتأكد من خلوها من مشاهدات شاذة قبل إجراء نموذج الانحدار.

- إن استخدام عدد كبير من المتغيرات الصورية كمتغيرات مفسرة في نموذج الانحدار يؤدي أيضاً إلى بروز مشكلة الارتباط الخطي. وتتفاقم المشكلة إذا تضمن النموذج متغيرات تفاعل (Interaction regressors) بين هذه المتغيرات. فمثلاً من (٤) متغيرات صورية يمكن تعريف (٦) متغيرات تفاعل بينها ليكون العدد الكلي للمتغيرات يساوي (١٠). عليه ينصح بتقليل عدد المتغيرات الصورية في النموذج ما أمكن، وذلك بدمج بعض فئات المتغير النوعي المتشابهة.

٧-٢-٦ مثال:

البيانات المستعرضة بالجدول رقم (٧-١) تختص بالواردات (Y) والناتج القومي الإجمالي (X_1)، كلها ببلايين الدولارات والرقم القياسي العام لأسعار المستهلكين (X_2)، للولايات المتحدة الأمريكية من العام ١٩٦٤ إلى ١٩٧٩م. المطلوب بناء نموذج انحدار الواردات على الناتج القومي والرقم القياسي للأسعار والكشف عن وجود ارتباط خطي بين المتغيرين المستقلين، واقترح حلاً مناسباً لمشكلة الارتباط الخطي المتعدد إن وجدت؟

الحل:

باستخدام طريقة المربعات الصغرى الاعتيادية تم الحصول على النموذج المقدّر التالي:

$$\hat{y}_i = -101.4885 + 0.078534361x_1 + 0.758554026x_2, \quad R^2 = 0.987366$$

$$\begin{array}{ccc} (33.0803) & (0.05596) & (0.76125) \\ (0.0090) & (0.1839) & (0.3372) \end{array}$$

حيث إن الأرقام بين الأقواس أسفل قيم معاملات الانحدار المقدرة هي الأخطاء المعيارية وقيم الاحتمال (p-value) المناظرة لكل معامل. وتشير النتائج إلى معنوية الانحدار ككل (P-value =0.00) وأن النموذج يفسر ٩٨,٧% (R^2=0.987) من التغير في الواردات خلال الفترة ١٩٦٤-١٩٧٩م.

جدول رقم (٧-١): الواردات والناتج القومي الإجمالي والرقم القياسي لأسعار المستهلكين للولايات المتحدة الأمريكية ١٩٦٤-١٩٧٩م

الرقم القياسي للأسعار (X_2)	الناتج القومي الإجمالي (X_1)	الواردات (Y)	العام
92.9	635.7	28.4	1964
94.5	688.1	32.0	1965
97.2	753.0	37.7	1966
100.0	796.3	40.6	1967
104.2	868.5	47.7	1968
109.8	935.5	52.9	1969
116.3	982.4	58.5	1970
121.3	1063.4	64.0	1971
125.3	1171.1	75.9	1972
133.1	1306.6	94.4	1973
147.7	1412.9	131.9	1974
161.2	1528.8	126.9	1975
170.5	1702.2	155.4	1976
181.5	1899.5	185.8	1977
195.4	2127.6	217.5	1978
217.4	2368.5	260.9	1979

المصدر: دومينيك سالفاتور (١٩٨٢) ص ٢١٠

الكشف عن وجود ارتباط خطي بين متغيري الناتج القومي والرقم القياسي للأسعار:

- على الرغم من معنوية الانحدار ككل وكبر حجم معامل التحديد ($R^2=0.9874$) إلا أن $\hat{\beta}_1$ و $\hat{\beta}_2$ ، معاملا الناتج القومي والرقم القياسي للأسعار، ليسا معنويين إحصائياً؛ حيث بلغت قيمتا الاحتمال ٠,١٨ و٠,٣٤ على التوالي، وهي إشارة واضحة لوجود ارتباط خطي بين الناتج القومي والرقم القياسي.
- الارتباط الخطي البسيط: يدعم وجود الارتباط الخطي بين الناتج القومي والسعر القياسي العلاقة شبه التامة بينهما، إذ بلغت قيمة معامل الارتباط الخطي البسيط ما يقارب الواحد صحيح ($r=0.997$).
- عامل تضخم التباين: لحساب عامل تضخم التباين يتم أولاً بناء نموذج انحدار الناتج القومي الإجمالي على الرقم القياسي للأسعار أو الرقم القياسي للأسعار على الناتج القومي الإجمالي للحصول على معامل التحديد. كما يمكن الحصول على معامل التحديد بتربيع قيمة معامل الارتباط البسيط بين الناتج المحلي الإجمالي والرقم القياسي للأسعار. وبما أن معامل تحديد نموذج انحدار الناتج القومي الإجمالي على الرقم القياسي للأسعار يساوي ٠,٩٩٤٣ ($R^2 = 0.9943$) فإن عامل تضخم التباين يتم حسابه حسب المعادلة (7.12) كما يلي:

$$VIF_1 = VIF_2 = \frac{1}{1-R^2_1} = \frac{1}{1-R^2_2} = \frac{1}{1-0.9943} = 176.64$$

وحيث إن قيمة عامل تضخم التباين أكبر بكثير من (١٠) -العتبة التي حددها العديد من الكُتاب لوجود الارتباط الخطي- نستنتج أن هناك ارتباطاً خطياً مرتفعاً جداً بين الناتج القومي الإجمالي والرقم القياسي للأسعار. وبما أن قيمة عامل التضخم لمعاملي الانحدار هي ١٧٦,٦٤ فإن قيمة متوسط عاملي تضخم التباين هي أيضاً ١٧٦,٦٤ وهذا يوضح أن مربع الخطأ في مقدرات المربعات الصغرى هو قرابة ١٧٧ مرة أكبر من قيمته في حال تعامد المتغيرات المستقلة. وهذا يوضح حجم مشكلة الارتباط الخطي الخطير التي يعاني منها هذا النموذج.

- **قيم الجذر الكامنة/المميزة:**

لحساب القيم الكامنة تتبع الخطوات التالية:

- من بيانات المثال تم حساب المصفوفة $\mathbf{Z}$ $\left(\mathbf{x}^T\mathbf{x}\right)$ التالية:

$$\mathbf{Z} = \left(\mathbf{x}^T\mathbf{x}\right) = \begin{pmatrix} 16 & 20240 & 2168 \\ 20240 & 29858334 & 3054767 \\ 2168 & 3054767 & 316834 \end{pmatrix}$$

ومن ثم تم حساب المصفوفة القطرية $\mathbf{S}$:

$$\mathbf{S}=\begin{pmatrix}0.25 & 0 & 0\\ 0 & 0.00018301 & 0\\ 0 & 0 & 0.00177658\end{pmatrix}$$

حيث إن قيمة أية عنصر من عناصر المصفوفة **S** عبارة عن معكوس الجذر التربيعي لعنصر المصفوفة **Z** المقابلة لها؛ أي إن:

$$\mathbf{S}_{ii}=\frac{1}{\sqrt{\mathbf{Z}_{ii}}}$$

وبضرب قبلي وبعدي للمصفوفة القطرية **S** في المصفوفة **Z** نتحصل على:

$$\mathbf{SZS}=\begin{pmatrix}0.25 & 0 & 0\\ 0 & 0.00018301 & 0\\ 0 & 0 & 0.00177658\end{pmatrix}\begin{pmatrix}16 & 20240 & 2168\\ 20240 & 29858334 & 3054767\\ 2168 & 3054767 & 316834\end{pmatrix}\begin{pmatrix}0.25 & 0 & 0\\ 0 & 0.00018301 & 0\\ 0 & 0 & 0.00177658\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0.92602 & 0.96304\\ 0.92602 & 1 & 0.99318\\ 0.96304 & 0.99318 & 1\end{pmatrix}$$

وباستخدام طرق المصفوفات لإيجاد القيم الكامنة/المميزة للمصفوفة **SZS** نتحصل على:

$$\lambda_1=2.92175,\quad \lambda_2=0.007799,\quad \lambda_3=0.0002569$$

والجدول التالي يوضح القيم الكامنة ومؤشرات الحالة للمتغيرات بما في ذلك المعامل الثابت. وبملاحظة القيم الكامنة نجد أن القيمة الكامنة المناظرة لمتغير الرقم القياسي صغيرة جداً (٠,٠٠٠٢٦) مما يشير إلى وجود الارتباط الخطي. أما رقم الحالة (Condition Number) – الجذر التربيعي لنسبة أعلى قيمة كامنة لأقل قيمة كامنة- البالغ قيمته (١٠٦,٦٥) فيدل بوضوح على وجود ارتباط خطي مرتفع جداً حيث تزيد قيمة المؤشر عن الحدود التي اقترحها كل من جونستون وبيلسلي وآخرين.

المتغير	القيم الكامنة (Eigenvalues)	مؤشر الحالة (Condition Index)
المعامل الثابت	2.92175	1
الناتج المحلي الإجمالي	0.07799	6.12075
الرقم القياسي للأسعار	0.0002569	106.65171

طرق المعالجة:

فيما سبق تعرضنا لعدد من الحلول الخاصة بمشكلة الارتباط الخطي. ومن بين الحلول المقترحة نجد أن زيادة عدد مشاهدات السلسلة الزمنية أو إجراء انحدار التل تمثلان أكثر الحلول ملائمة. أما الحلول الأخرى تصلح في حالات تختلف عن هذا المثال. فإسقاط أي من المتغيرين المستقلين يؤدي إلى تحيز في التقدير وإجراء تحليل المكونات الأساسية أو التحليل العاملي يستخدم في حالة وجود متغيرات متعددة؛ أكثر من متغيرين. فيما يلي نستخدم طريقة انحدار التل لتقدير معالم نموذج انحدار الواردات على الناتج المحلي الإجمالي والأرقام القياسية للأسعار.

مقدرات انحدار التل:

لبناء نموذج انحدار التل يتم اختيار قيم تحيزية مختلفة وإجراء حل النموذج لكل قيمة من هذه القيم. ولتحديد قيمة التحيز c التي تعطي أفضل نموذج، يستخدم عادةً الرسم البياني الخطي لقيم معاملات انحدار التل (المحور الصادي) مع قيم مختلفة لثابت التحيز ذات مسافات متساوية (المحور الأفقي). كما يؤخذ في الاعتبار قيمة عامل تضخم التباين للتأكد من حل مشكلة الارتباط الخطي. فإذا أظهر الشكل استقراراً في قيم معاملات الانحدار وانخفاض قيم عوامل تضخم التباين عند قيم محددة لثابت التحيز، يتم اختيار أحد النماذج المناظرة بصورة تحكمية. ولتوضيح كيفية تقدير معاملات انحدار التل المعياري، نقوم بإجراء النموذج باستعمال ثابت تحيز مساو لـ ٠,٥. وفيما يلي خطوات حل النموذج:

-مصفوفة الارتباط الخطي البسيط بين المتغيرات المستقلة:

$$\mathbf{r}_{xx} = \begin{pmatrix} 1.0 & 0.99717 \\ 0.99717 & 1.0 \end{pmatrix}$$

-متجه معاملات الارتباط الخطي البسيط بين المتغير التابع مع المتغيرين المستقلين:

$$\mathbf{r}_{yx} = \begin{pmatrix} 0.993177 \\ 0.992699 \end{pmatrix}$$

- مصفوفة وحدة من درجة 2x2:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- وبضرب ثابت التحيز c=0.5 في مصفوفة الوحدة نتحصل على:

$$c\mathbf{I} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

- وبجمع المصفوفة $\mathbf{r}_{xx}$ للمصفوفة $c\mathbf{I}$ نتحصل على:

$$\mathbf{r}_{xx} + c\mathbf{I} = \begin{pmatrix} 1.5 & 0.99717 \\ 0.99717 & 1.5 \end{pmatrix}$$

- وباستعمال طرق المصفوفات لإيجاد معكوس المصفوفة $(\mathbf{r}_{xx} + c\mathbf{I})^{-1}$ نتحصل على:

$$(\mathbf{r}_{xx} + c\mathbf{I})^{-1} = \begin{pmatrix} 1.19459 & -0.79414 \\ -0.79414 & 1.19459 \end{pmatrix}$$

- وبضرب $(\mathbf{r}_{xx} + c\mathbf{I})^{-1}$ في $\mathbf{r}_{xy}$ نتحصل على:

$$(\mathbf{r}_{xx} + c\mathbf{I})^{-1}\mathbf{r}_{yx} = \begin{pmatrix} 1.19459 & -0.79414 \\ -0.79414 & 1.19459 \end{pmatrix}\begin{pmatrix} 0.993177 \\ 0.992699 \end{pmatrix} = \begin{pmatrix} 0.398102 \\ 0.397151 \end{pmatrix}$$

وعليه فإن تقدير انحدار التل المعياري لمعالم النموذج هو:

$$\hat{\beta}_1^R = 0.398102 \qquad \hat{\beta}_2^R = 0.397151$$

ويعطي الجدول رقم (٧-٢) قيم معاملات نموذج انحدار الواردات المقدرة باستخدام طريقة انحدار التل المعياري المناظرة لقيم مختلفة لثابت التحيز. ويتضح من الشكل رقم (٧-٢) إن قيم معاملي الانحدار غير مستقرة بالنسبة لقيم ثابت التحيز c الصغيرة. ويلاحظ أن الاستقرار في معاملي الانحدار عند قيم c التي تتراوح ما بين 0.05 إلى 0.09 . كما يتضح من الشكل رقم (٧-٣) أن عوامل تضخم التباين (VIF's) قد أخذت قيماً أقل من الواحد الصحيح عند قيمة c=0.05 فأكبر. كما يجب ملاحظة أنه من المرغوب فيه اختيار أصغر قيمة لـ c التي يحدث عندها الاستقرار طالما أن قيمة c مرتبطة مباشرة بقيمة التحيز الناتج.

ووفقاً للمعادلة (7.25) يتم حساب قيمة ثابت التحيز المثلى كما يلي:

$$c = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}^{R^T}\hat{\boldsymbol{\beta}}^R} = \frac{2 \times 0.014578}{\begin{pmatrix} 0.581486 & 0.412861 \end{pmatrix}\begin{pmatrix} 0.581486 \\ 0.412861 \end{pmatrix}} = 0.057$$

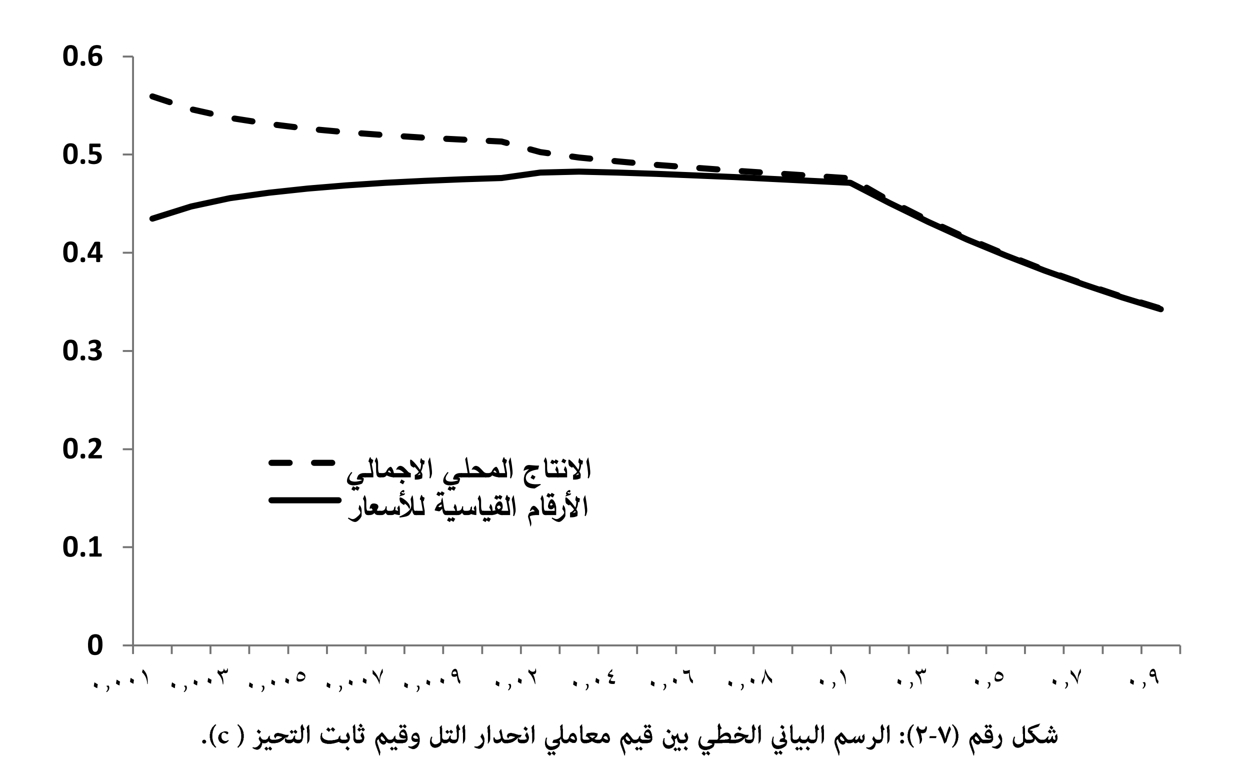
ويلخص الجدول رقم (٧-٣) نتائج انحدار التل عند قيمة ثابت التحيز (c=0.057) مقارنة بنتائج طريقة المربعات الصغرى الاعتيادية.

جدول رقم (٧-٢): مقدرات انحدار التل لنموذج الوردات على الناتج المحلي الإجمالي والأرقام القياسية للأسعار لقيم مختلفة لثابت التحيز c

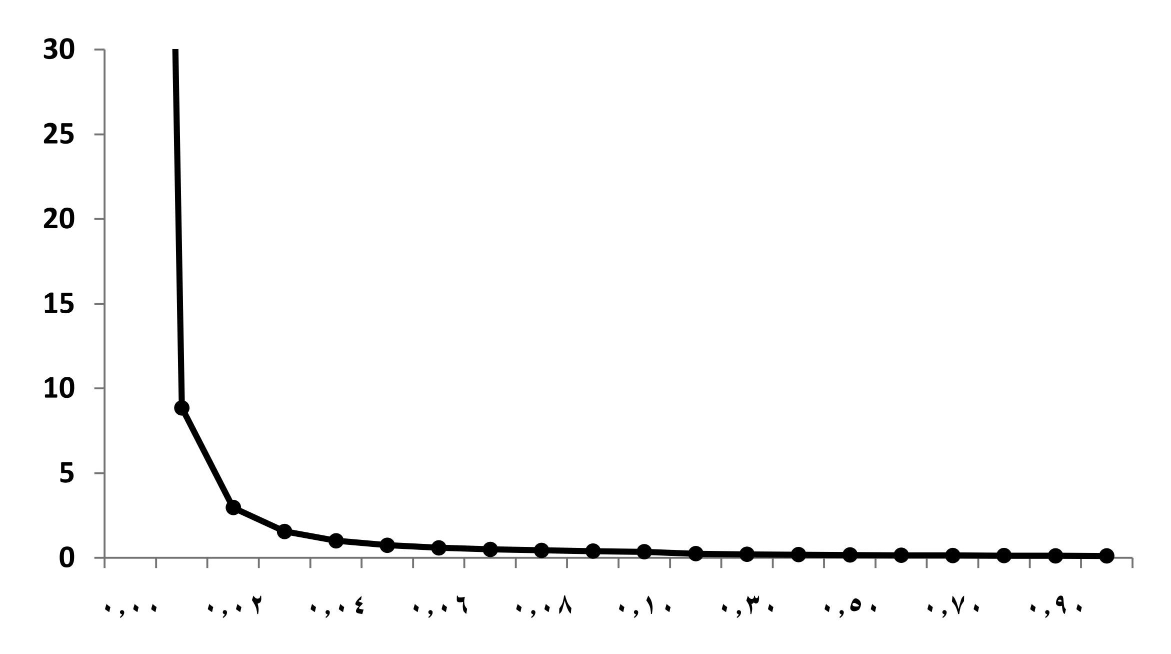
c	$\hat{\beta}_1^R$	$\hat{\beta}_2^R$	VIF	R^2
0.0000	0.581486	0.412861	176.64	98.74
0.0010	0.559251	0.434599	96.64	98.69
0.0020	0.546111	0.447242	60.89	98.64
0.0030	0.53739	0.455466	41.88	98.59
0.0040	0.531148	0.461212	30.59	98.54
0.0050	0.526437	0.465427	23.34	98.49
0.0060	0.522737	0.468632	18.41	98.44
0.0070	0.519739	0.471136	14.90	98.39
0.0080	0.517249	0.473132	12.32	98.34
0.0090	0.515138	0.474749	10.37	98.29
0.0100	0.513318	0.476075	8.85	98.24
0.0200	0.502711	0.481778	2.96	97.75
0.0300	0.497095	0.482537	1.56	97.27
0.0400	0.492991	0.481832	1.01	96.79
0.0500	0.489554	0.480507	0.75	96.32
0.0600	0.486477	0.478869	0.59	95.85
0.0700	0.483619	0.477057	0.50	95.39
0.0800	0.480911	0.47514	0.44	94.93
0.0900	0.47831	0.473161	0.39	94.48
0.1000	0.475791	0.471143	0.36	94.02
0.2000	0.453096	0.45074	0.24	89.75
0.3000	0.433034	0.431456	0.20	85.84
0.4000	0.414807	0.41362	0.18	82.26
0.5000	0.398101	0.397151	0.17	78.96
0.6000	0.382713	0.38192	0.15	75.92
0.7000	0.368481	0.367801	0.14	73.11
0.8000	0.355278	0.354682	0.13	70.49
0.9000	0.342992	0.342463	0.12	68.06
1.0000	0.331531	0.331054	0.11	65.79

جدول رقم (٧-٣): مقارنة نتائج طريقة المربعات الصغرى الاعتيادية مع نتائج انحدار التل عند قيمة ثابت التحيز (c=0.057)

المتغير	طريقة المربعات الصغرى C=0.0 $\hat{\beta}^R$	طريقة المربعات الصغرى C=0.0 $\hat{\beta}$	طريقة انحدار التل C=0.057 $\hat{\beta}^R$	طريقة انحدار التل C=0.057 $\hat{\beta}$
المعامل الثابت	-	-101.4885	-	-101.973
الناتج المحلي الإجمالي	0.581	0.07853	0.487372	0.06582
الأرقام القياسية للأسعار	0.413	0.75855	0.479384	0.88078
معامل التحديد (r^2)	0.987366		0.959931	
عامل تضخم التباين	176.64		0.632535	



شكل رقم (٧-٢): الرسم البياني الخطي بين قيم معاملي انحدار التل وقيم ثابت التحيز (c).



شكل رقم (٧-٣): الرسم البياني الخطي لقيم عامل تضخم التباين عند قيم مختلفة لثابت التحيز (c).

٧-٣ اختلاف التباين (Heteroscedasticity):

٧-٣-١ مقدمة:

من الاشتراطات الأساسية التي تقوم عليها طريقة المربعات الصغرى لتقدير معالم نموذج الانحدار الخطي هو ثبات تباين حد الخطأ عند كل مستوى من مستويات المتغير أو المتغيرات المستقلة. ويعرف هذا الاشتراط بثبات التباين (Homoscedasticity)، أي أن:

$$\text{var}(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 \qquad i = 1,2,\ldots,N$$

وباستخدام رموز المصفوفات نجد أن:

$$E(\boldsymbol{\varepsilon\varepsilon}^T) = \sigma^2 \mathbf{I}_{NXN} = \begin{pmatrix} \sigma^2 & 0 & 0 & . & . & . & 0 \\ 0 & \sigma^2 & 0 & . & . & . & 0 \\ 0 & 0 & \sigma^2 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & \sigma^2 \end{pmatrix} \qquad (7\text{-}28)$$

ويلاحظ أن العنصر القطري لهذه المصفوفة يحتوي على قيم ثابتة (σ^2) وأن جميع العناصر خارج القطر هي قيم صفرية. وبعدم استيفاء هذا الشرط نواجه بمشكلة تعرف باختلاف التباين أو عدم ثبات التباين (Heteroscedasticity)، حيث تكون تباينات حدود الخطأ مختلفة كما في الشكل رقم (٧-٤) ، أو ما يمكن التعبير عنه رياضياً كالآتي:

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)=\begin{pmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ 0 & 0 & \sigma_3^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \sigma_N^2 \end{pmatrix} \qquad (7\text{-}29)$$

حيث إن $\sigma_1^2,\sigma_2^2,\ldots,\sigma_N^2$ قيم مختلفة. وباستخدام العلاقة التالية:

$$\sigma_i^2=\sigma^2 W_i$$

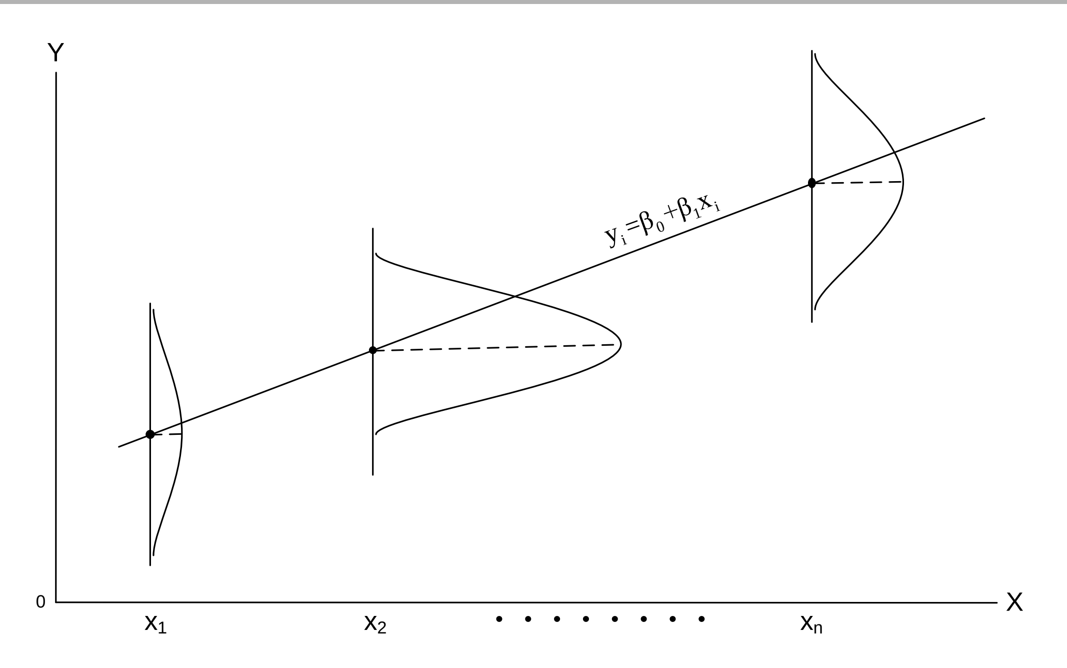
حيث إن W_i أوزان تأخذ قيماً مختلفة. ويمكن كتابة المصفوفة (7.29) كما يلي:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)=\sigma^2\mathbf{W}$$

حيث أن W مصفوفة قطرية من الدرجة NxN يحتوي قطرها الرئيسي على القيم w_i وجميع عناصرها خارج القطر مساوية للصفر، أي أن:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)=\sigma^2\begin{pmatrix} w_1 & 0 & 0 & . & . & . & 0 \\ 0 & w_2 & 0 & . & . & . & 0 \\ 0 & 0 & w_3 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & w_N \end{pmatrix} \qquad (7\text{-}30)$$

حيث إن $w_1,w_2,\ldots,w_N$ قيم مختلفة.



شكل رقم (٧-٤): توزيع المتغير العشوائي (ε_i)، حالة عدم ثبات التباين

٧-٣-٢ أسباب عدم ثبات التباين:

تحدث مشكلة عدم ثبات التباين في البيانات المقطعية (Cross-sectional data) أكثر من بيانات السلاسل الزمنية (Time series data). وفيما يلي بعض الحالات التي قد نواجه فيها هذه المشكلة:

- مع زيادة دخول الأفراد يزداد تباين إنفاقهم، حيث يلاحظ أن الأفراد ذوي الدخول المنخفضة يكون تباين إنفاقهم منخفضاً وذلك لاقتصار إنفاقهم على الضروريات، في حين نجد أن تباين الإنفاق يكون كبيراً عند مستويات الدخول المرتفعة نظراً لتنوع أوجه الإنفاق على السلع غير الضرورية. ولذا يتوقع عند بناء نموذج انحدار الإنفاق على الدخل أن يزداد تباين حد الخطأ بزيادة الدخل.

- وقد نواجه أيضاً بمشكلة عدم ثبات التباين عندما يكون هناك خطأ في قياس المتغير التابع ويختلف حجم هذا الخطأ باختلاف قيم المتغير المستقل. فمثلاً يتوقع الحصول على معلومات دقيقة في التعداد السكاني من سكان المد ن مقارنة بالمعلومات التي يمكن الحصول عليها من سكان الريف نظراً لاختلاف الظروف الاقتصادية والاجتماعية والثقافية.

- وجود أخطاء في توصيف نموذج الانحدار، وذلك إما بأن يكون الشكل الرياضي للنموذج غير ملائم أو لم يتم إدراج بعض المتغيرات المفسرة.

٧-٣-٣ النتائج المترتبة على وجود اختلاف التباين:

باستيفاء فروض طريقة المربعات الصغرى نجد أن $\widehat{\boldsymbol{\beta}}$ أفضل مقدر خطي غير متحيز (BLUE) لمعالم نموذج الانحدار. وفي وجود اختلاف التباين يمكن تقدير معالم نموذج الانحدار الخطي التي ستظل غير متحيزة ومتسقة ولكنها تفقد خاصية الكفاءة، أي خاصية أقل تباين.

القيمة المتوقعة وتباين معالم نموذج الانحدار المقدرة في ظل عدم ثبات التباين:

في ظل عدم ثبات التباين يمكن الحصول على مقدرات المربعات الصغرى والتي تحتفظ بخاصية عدم التحيز. ويمكن برهان ذلك فيما يلي:

بما أن نموذج الانحدار المتعدد يأخذ الصيغة التالية:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

وأن

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y}$$

فإن

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}(\mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{I}\boldsymbol{\beta} + \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\boldsymbol{\varepsilon}$$

وبأخذ القيمة المتوقعة لطرفي المعادلة نجد أن:

$$\mathrm{E}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$$

إلا أن تباين معالم نموذج الانحدار المقدرة لم يعد يمكن وفق المعادلة التالية (انظر الفصل الثالث):

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}$$

وفي ظل اختلاف التباين نجد أن تباين مقدرات المربعات الصغرى يتخذ الصيغة التالية:

$$\begin{aligned}
\mathrm{var}(\widehat{\boldsymbol{\beta}}) &= \mathrm{E}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}} \\
&= \mathrm{E}\{(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{x}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\} \\
&= (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\mathbf{x}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1} \\
&= \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{w}\mathbf{x}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}
\end{aligned}$$

وتوضح هذه المعادلة أن تباين معالم نموذج الانحدار المقدرة في حالة عدم ثبات التباين يختلف عن تباينها في حالة استيفاء فرضية ثبات التباين. ويترتب على هذا الاختلاف أن تكون النتائج التي نحصل عليها من الاستدلال الإحصائي – اختبارات المعنوية وفترات الثقة وفترات التنبؤ- غير صحيحة.

٧-٣-٤ بعض الطرق المستخدمة للكشف عن اختلاف التباين:

تعرضنا في الفصل الثالث لبعض الطرق البيانية التي تساعد في الكشف عن عدم ثبات التباين باستخدام تحليل البواقي. وفي هذا الجزء سندرس بعضاً من الطرق التحليلية المستخدمة للكشف عن وجود هذه المشكلة.

٧-٣-٤-١ اختبار بارك (Park, 1969):

يقترح بارك العلاقة التالية بين تباين حد الخطأ (σ_i^2) والمتغير المستقل (X_i):

$$\sigma_i^2 = \sigma^2 x_i^{\beta_1} e^{v_i} \qquad (7\text{-}31)$$

وبأخذ لوغاريتم طرفي المعادلة نحصل على:

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta_1 \ln x_i + v_i \qquad (7\text{-}32)$$

حيث إن V_i متغير عشوائي. وبما أن قيمة σ_i^2 غير معلومة، يقترح بارك استخدام مربع البواقي (e_i^2) كمقرب (Proxy) وإجراء نموذج الانحدار التالي:

$$\ln e_i^2 = \ln\sigma^2 + \beta_1 \ln x_i + v_i$$

أو

$$\ln e_i^2 = \beta_0 + \beta_1 \ln x_i + v_i \qquad (7\text{-}33)$$

حيث إن $\beta_0 = \ln \sigma^2$

فإذا كانت β_1 معنوية إحصائياً دلّ ذلك على وجود مشكلة عدم ثبات التباين. وأما إذا كانت β_1 غير دالة إحصائياً، فإنه يمكن القول بأن تباين حد الخطأ ثابت.

ويلاحظ الآتي على اختبار بارك:

- يتضمن اختبار بارك خطوتين هما: الخطوة الأولى؛ وفيها يتم بناء نموذج الانحدار والحصول على البواقي، وفي الخطوة الثانية يتم إجراء انحدار لوغاريتم مربع البواقي على لوغاريتم المتغير المستقل حسب المعادلة (7.33).
- يعاب على اختبار بارك أن المتغير العشوائي v_i في المعادلة (7.32) أو (7.33) قد لا يستوفي فرضيات المربعات الصغرى والتي من ضمنها فرضية ثبات التباين (Gujarati and Porter, 2009, p.379).
- يستخدم اختبار بارك للكشف عن عدم ثبات التباين في نموذجي الانحدار الخطي البسيط والانحدار الخطي المتعدد. وفي حالة نموذج الانحدار الخطي المتعدد يتم حساب البواقي ويجرى انحدار لوغاريتم مربع البواقي مع المتغير المستقل (X_j) مصدر اختلاف التباين والذي يتغير معه التباين.

٧-٣-٤-٢ اختبار جليجسر (Glejser; 1969):

يعد اختبار جليجسر شبيهاً باختبار بارك، فبعد الحصول على البواقي (e_i) من نموذج الانحدار الأصلي، يقترح جليجسر إجراء انحدار القيم المطلقة للبواقي ($|e_i|$) على المتغير المستقل (X_i) الذي يتغير معه التباين (σ_i^2). وقد استخدم جليجسر الأشكال الدالية التالية:

$$|e_i| = \beta_0 + \beta_1 x_i + v_i \qquad (7\text{-}34)$$

$$|e_i| = \beta_0 + \beta_1 \sqrt{x_i} + v_i \qquad (7\text{-}35)$$

$$|e_i| = \beta_0 + \beta_1 \frac{1}{x_i} + v_i \qquad (7\text{-}36)$$

$$|e_i| = \beta_0 + \beta_1 \frac{1}{\sqrt{x_i}} + v_i \qquad (7\text{-}37)$$

$$|e_i| = \sqrt{\beta_0 + \beta_1 x_i} + v_i \qquad (7\text{-}38)$$

$$|e_i| = \sqrt{\beta_0 + \beta_1 x_i^2} + v_i \qquad (7\text{-}39)$$

حيث إن v_i حد الخطأ العشوائي.

فإذا كانت β_1 معنوية إحصائياً دلّ ذلك على وجود مشكلة عدم ثبات التباين. وأما إذا كانت β_1 غير معنوية إحصائياً فإن النموذج يستوفي شرط ثبات التباين. ويعاب على طريقة جليجسر أن القيمة المتوقعة للحد العشوائي قد تختلف عن الصفر، وكذلك قد يكون تباين حد الخطأ غير ثابت فضلاً عن أن المعادلتين (7.38) و(7.39) غير خطية المعالم الأمر الذي يجعل استخدام طريقة المربعات الصغرى الاعتيادية لتقدير معالمها مستحيلاً. هذا وقد وجد جليجسر أن المعادلات الأربعة الأولى (7.34) إلى (7.37) تحقق نتائج جيّدة يعتد بها في الكشف عن عدم ثبات التباين.

٧-٣-٤-٣ اختبار جولدفيلد-كواندت (Goldfeld-Quandt, 1965):

يعد اختبار جولدفيلد-كوندت من الاختبارات الشائعة الاستخدام في الكشف عن وجود علاقة طردية أو عكسية بين تباين حد الخطأ وأحد المتغيرات المستقلة. ويتلخص اختبار جولدفيلد-كواندت في الخطوات التالية:

- ترتيب المشاهدات حسب قيم المتغير المستقل (X_i) تصاعدياً، وفي حالة نموذج الانحدار المتعدد يتم الترتيب حسب قيم المتغير المستقل مصدر اختلاف التباين.
- يتم استبعاد عدد (d) مشاهدة من الوسط – في حدود ربع عدد المشاهدات- ومن ثم يتم تقسيم بقية متغيرات المشاهدات إلى مجموعتين بحيث يكون عدد كل مجموعة يساوي $(n-d)/2$ مشاهدة.
- يتم بناء نموذجين، أحدهما للقيم الصغيرة للمتغير (X) والآخر للقيم الكبيرة ويتم الحصول على مجموع مربعات البواقي للنموذجين، RSS_1 وRSS_2 على التوالي.

- يستخدم اختبار جولدفيلد-كواندت للكشف عن نوعين من عدم ثبات التباين هما:

- النوع الأول: تباين حد الخطأ دالة تزايدية للمتغير المستقل (X_i) والفرض المراد اختباره هو:

فرض العدم: تباين حد الخطأ ثابت مقابل **الفرض البديل**: تباين حد الخطأ دالة تزايدية للمتغير X_i.

ولإجراء هذا الاختبار تستخدم الإحصاءة التالية:

$$F_I = \frac{RSS_2/V_2}{RSS_1/V_1} \quad \sim F_{V_2,V_1} \qquad (7\text{-}40)$$

حيث إن:

RSS_1 = مجموع مربعات البواقي لنموذج مشاهدات المجموعة الأولى (القيم الصغيرة لـ X_i).

RSS_2 = مجموع مربعات البواقي لنموذج مشاهدات المجموعة الثانية (القيم الكبيرة لـ X_i).

p = عدد المتغيرات المستقلة.

و

$$V_1 = V_2 = \frac{(n-d)}{2} - p - 1$$

وتتوزع الإحصاءة F_I حسب توزيع F بدرجتي حرية V_2 و V_1. فإذا كانت قيمة F_I المحسوبة أكبر من قيمة F الجدولية فإننا نرفض فرض العدم ويتم قبول الفرض البديل القائل بأن تباين حد الخطأ يتزايد بزيادة قيم المتغير المستقل؛ وأما إذا كانت قيمة F_I المحسوبة أقل من قيمة F الجدولية فيتم قبول فرض العدم الذي ينص على أن تباين حد الخطأ ثابت.

- النوع الثاني: تباين حد الخطأ دالة تناقصية للمتغير المستقل (X_i) والفرض المراد اختباره هو:

فرض العدم: تباين حد الخطأ متجانس أو ثابت في مقابل **الفرض البديل:** تباين حد الخطأ دالة تناقصية للمتغير X_i.

ولإجراء هذا الاختبار تستخدم الإحصاءة التالية:

$$F_{II} = \frac{RSS_1/V_1}{RSS_2/V_2} \quad \sim F_{V_1,V_2} \qquad (7\text{-}41)$$

وتتوزع الإحصاءة F_{II} حسب توزيع F بدرجتي حرية V_1 و V_2. فإذا كانت قيمة F_{II} المحسوبة أكبر من قيمة F الجدولية، فإننا نرفض فرض العدم ويتم قبول الفرض البديل القائل بأن تباين حد الخطأ يتناقص بزيادة قيم المتغير المستقل؛ وأما إذا كانت قيمة F_{II} المحسوبة أقل من قيمة F الجدولية فيتم قبول فرض العدم الذي ينص على أن تباين حد الخطأ متجانس.

٧-٣-٤- ٤ اختبار بروش-باقان (Breusch & Pagan Test):

يعد اختبار بروش-باقان (Breusch & Pagan, 1979) من الاختبارات العامة التي تستخدم للكشف عن وجود عدد كبير من حالات مشكلة عدم ثبات التباين. ويعتمد هذا الاختبار على بواقي نموذج الانحدار المقدر باستخدام طريقة المربعات الصغرى الاعتيادية. فمن نموذج الانحدار الخطي:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

حيث إن $\boldsymbol{\varepsilon}$ متجه حدود الخطأ يفترض أن يتبع التوزيع الطبيعي وأن تكون حدود الخطأ مستقلة ولها تباين

$$\sigma_i^2 = f\,(\theta_1 + \theta_2 Z_{2i} + \theta_3 Z_{3i} + ... + \theta_m Z_{mi}) \qquad (7\text{-}42)$$

حيث إن:

$f(.)$ دالة غير محددة الشكل.

$\theta_1\ \theta_2\ \theta_3\ ...\theta_m$ معاملات لا علاقة لها بمعالم نموذج الانحدار الأصلي ($\boldsymbol{\beta}$) .

$Z_2\ Z_3\ ...\ Z_m$ المتغيرات التي يعتقد أنها سبب اختلاف التباين.

ولاختبار ثبات التباين يستخدم فرض العدم التالي: $H_0 : \theta_2 = \theta_3 = ... = \theta_m = 0$ فإذا لم يتم رفض فرض العدم، يعني ثبات التباين، أي $\sigma_i^2 = \theta_1$.

ولإجراء هذا الاختبار يتم اتباع الخطوات التالية:

- تقدير نموذج الانحدار الأصلي باستخدام طريقة المربعات الصغرى الاعتيادية ويتم الحصول على بواقي النموذج الموفق (e_i).

- ومن البواقي يتم حساب التباين المتحيز (مقدر طريقة الإمكان الأعظم (MLE)) التالي:

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n}$$

ومن ثم يتم حساب المتغير الجديد التالي:

$$g_i = \frac{e_i^2}{\tilde{\sigma}^2} \qquad i = 1,2,...,n \qquad (7\text{-}43)$$

- يتم تحديد المتغيرات التي يعتقد أنها سبب عدم ثبات التباين والتي تشمل بعض أو كل المتغيرات المستقلة في النموذج الأصلي ($Z_2\ Z_3\ ...\ Z_m$)، ومن ثم إجراء انحدار g_i على هذه المتغيرات، أي:

$$g_i = \theta_1 + \theta_2 Z_{2i} + \theta_3 Z_{3i} + ... + \theta_m Z_{mi} + v_i$$

- يتم حساب مجموع مربعات الانحدار (ESS) من نموذج الانحدار الموفق.
- وفي حالة العينات الكبيرة نجد أن نصف مجموع مربعات الانحدار (ESS/2) تحت ظل فرض العدم يتبع توزيع مربع كاي عند درجات حرية (p)-عدد المتغيرات المستقلة- أي أن:

$$Q = \frac{ESS}{2} \sim \chi^2_m$$

- في هذه الخطوة يتم إجراء الاختبار الإحصائي بمقارنة قيمة Q بالقيمة الجدولية لتوزيع مربع كاي عند درجة حرية (p) ومستوى معنوية معين. فإذا كانت

$$Q = \frac{ESS}{2} > \chi^2_{\alpha,m} \qquad (7\text{-}44)$$

نرفض فرض ثبات التباين.

٧-٣-٤- ٥ اختبار وايت (White's Test):

طور وايت (White, 1980) اختباراً مباشراً للكشف عن عدم ثبات التباين، قريبٌ جداً من حيث صياغته الدالية لاختبار بروش-باقان. ويتميز بأنه سهل التطبيق ولا يعتمد على شرط الاعتدالية أو التوزيع الطبيعي. وفيما يلي خطوات اختبار بروش – باقان:

- تقدير نموذج الانحدار الأصلي باستخدام طريقة المربعات الصغرى الاعتيادية ويتم الحصول على بواقي النموذج الموفق (e_i).
- يتم بناء نموذج انحدار e_i^2 على المتغيرات المستقلة في النموذج الأصلي ($Z_2 \; Z_3 \; ... \; Z_m$) بالإضافة إلى متغيرات نواتج حاصل ضرب المتغيرات المستقلة مع بعضها ($Z_2^2, Z_3^2, \; Z_2Z_3...$)، أي:

$$e_i^2 = \theta_1 + \theta_2 Z_{2i} + \theta_3 Z_{3i} + \theta_4 Z_{2i}^2 + \theta_5 Z_{3i}^2 + \theta_6 Z_{2i} Z_{3i} + ... + \theta_m Z_{mi} + v_i$$

- يتم حساب معامل التحديد من نموذج الانحدار الموفق R^2 .
- وفي حالة العينات الكبيرة نجد أن حاصل ضرب عدد المشاهدات في معامل التحديد (nR^2) تحت ظل فرض العدم يتبع توزيع مربع كاي عند درجات حرية (m)-عدد المتغيرات المستقلة- أي أن:

$$Q = n \times R^2 \sim \chi^2_m$$

- في هذه الخطوة يتم إجراء الاختبار الإحصائي بمقارنة قيمة Q بالقيمة الجدولية لتوزيع مربع كاي عند درجة حرية (m) ومستوى معنوية معين. فإذا كانت

$$Q = n \times R^2 > \chi^2_{\alpha,m} \qquad (7\text{-}45)$$

نرفض فرض ثبات التباين.

٧-٣-٥ بعض طرق معالجة مشكلة عدم ثبات التباين:

كما أسلفنا فإنه يمكن الحصول على مقدرات غير متحيزة ومتسقة في ظل عدم ثبات تباين حد الخطأ إلا أنها تصبح غير كفؤة. وحسب جون فوكس (Fox (1997) p.305) إن طريقة المربعات الصغرى تقل كفاءتها إذا كانت نسبة أكبر تباين لأقل تباين لحد الخطأ تساوى (١٠) أو أكثر. ويترتب على اختلال خاصية الكفاءة أن تكون قيم الأخطاء المعيارية لمعاملات الانحدار المقدرة غير صحيحة. وفيما يلي نستعرض بعض الطرق المستخدمة في علاج المشكلة.

- **استخدام طريقة المربعات الصغرى المرجحة عندما تكون قيم σ_i^2 معلومة:**

نعلم أن صيغة نموذج الانحدار الخطي المتعدد هي:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

وأن تباين حد الخطأ في ظل عدم ثبات التباين يتخذ الصيغة التالية:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{W}$$

ولعلاج مشكلة عدم ثبات التباين يتم تصحيح نموذج الانحدار الخطي المتعدد وذلك بضرب قبلي لطرفي المعادلة بالمصفوفة $\mathbf{W}^{-\frac{1}{2}}$ كالآتي:

$$\mathbf{W}^{-\frac{1}{2}}\mathbf{y} = \mathbf{W}^{-\frac{1}{2}}\mathbf{x}\boldsymbol{\beta} + \mathbf{W}^{-\frac{1}{2}}\boldsymbol{\varepsilon}$$

أو

$$\mathbf{y}^* = \mathbf{x}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \qquad (7\text{-}46)$$

حيث إن $\mathbf{y}^* = \mathbf{W}^{-\frac{1}{2}}\mathbf{y}$ ، $\mathbf{x}^* = \mathbf{W}^{-\frac{1}{2}}\mathbf{x}$ ، $\boldsymbol{\varepsilon}^* = \mathbf{W}^{-\frac{1}{2}}\boldsymbol{\varepsilon}$.

$$\mathbf{W}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \ldots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \frac{1}{\sigma_N} \end{pmatrix}$$

ويلاحظ أن هذه الطريقة عبارة عن تحويل المتغيرات المستقلة والمتغير التابع وذلك بقسمتها على الانحراف المعياري لحد الخطأ، ويمكن تبيان ذلك على النحو الآتي:

$$\mathbf{y}^* = \mathbf{W}^{-\frac{1}{2}}\mathbf{y} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_N} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_N \end{pmatrix} = \begin{pmatrix} \frac{y_1}{\sigma_1} \\ \frac{y_2}{\sigma_2} \\ \frac{y_3}{\sigma_3} \\ . \\ \frac{y_N}{\sigma_N} \end{pmatrix}$$

و

$$\mathbf{x}^* = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_N} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ 1 & x_{13} & x_{23} & \dots & x_{p3} \\ . & . & . & \dots & . \\ 1 & x_{1N} & x_{2N} & \dots & x_{pN} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1} & \frac{x_{11}}{\sigma_1} & \frac{x_{21}}{\sigma_1} & \dots & \frac{x_{p1}}{\sigma_1} \\ \frac{1}{\sigma_2} & \frac{x_{12}}{\sigma_2} & \frac{x_{22}}{\sigma_2} & \dots & \frac{x_{p2}}{\sigma_2} \\ \frac{1}{\sigma_3} & \frac{x_{13}}{\sigma_3} & \frac{x_{23}}{\sigma_3} & \dots & \frac{x_{p3}}{\sigma_3} \\ . & . & . & \dots & . \\ \frac{1}{\sigma_N} & \frac{x_{1N}}{\sigma_N} & \frac{x_{2N}}{\sigma_N} & \dots & \frac{x_{pN}}{\sigma_N} \end{pmatrix}$$

و

$$\boldsymbol{\varepsilon}^* = W^{-\frac{1}{2}}\boldsymbol{\varepsilon} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_N} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ \varepsilon_N \end{pmatrix} == \begin{pmatrix} \frac{\varepsilon_1}{\sigma_1} \\ \frac{\varepsilon_2}{\sigma_2} \\ \frac{\varepsilon_3}{\sigma_3} \\ . \\ \frac{\varepsilon_N}{\sigma_N} \end{pmatrix}$$

وبعد ترجيح المتغيرات تستخدم طريقة المربعات الصغرى الاعتيادية (غير المرجحة) على النحو التالي:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{x}^{*\mathrm{T}}\mathbf{x}^*)^{-1}\mathbf{x}^{*\mathrm{T}}\mathbf{y}^* \\ &= \left(\mathbf{x}^{\mathrm{T}}\mathbf{W}^{-\frac{1}{2}\mathrm{T}}\mathbf{W}^{-\frac{1}{2}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{W}^{-\frac{1}{2}\mathrm{T}}\mathbf{W}^{-\frac{1}{2}}\mathbf{y} \\ &= \left(\mathbf{x}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{y} \end{aligned}$$

ويتم الحصول على تباين معاملات نموذج الانحدار المقدرة كما يلي:

$$\mathrm{var}(\widehat{\beta}) = \sigma^2(\mathbf{x}^{*\mathrm{T}}\mathbf{x}^*)^{-1} = \sigma^2(\mathbf{x}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{x})^{-1}$$

حيث إن:

$$\mathbf{W}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sigma_3^2} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_N^2} \end{pmatrix}$$

ويلاحظ أن حد الخطأ الجديد يستوفي اشـتراطات طريقـة المربعـات الصـغرى العاديـة التـي مـن بينهـا شرط ثبـات التباين، ويمكن برهان ذلك على النحو التالي:

القيمة المتوقعة:

$$\mathrm{E}(\boldsymbol{\varepsilon}^*) = \mathrm{E}\left(\mathbf{W}^{-\frac{1}{2}}\boldsymbol{\varepsilon}\right) = \mathbf{0}$$

التباين- التغاير:

$$\begin{aligned}\mathrm{E}(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\mathrm{T}}) &= \mathrm{E}(\mathbf{W}^{-\frac{1}{2}}\varepsilon\varepsilon^{\mathrm{T}}\mathbf{W}^{-\frac{1}{2}\mathrm{T}}) \\ &= \mathbf{W}^{-\frac{1}{2}}\mathrm{E}(\varepsilon\varepsilon^{\mathrm{T}})\mathbf{W}^{-\frac{1}{2}\mathrm{T}} \\ &= \mathbf{W}^{-\frac{1}{2}}\sigma^2\mathbf{W}\mathbf{W}^{-\frac{1}{2}\mathrm{T}} \\ &= \sigma^2\mathbf{I}_{\mathrm{N}\times\mathrm{N}}\end{aligned}$$

- **استخدام طريقة المربعات الصغرى المرجحة عندما تكون قيم σ_i^2 مجهولة:**

في الواقع العملي نادراً ما تكون قيم σ_i^2 معلومة ولذلك لا بد من تقديرها. ويمكن أن يستخدم تباين في هذه الحالة بواقي نموذج الانحدار الأصلي كأوزان لتصحيح النموذج نفسه. حيث يتم استبدال قـيم الأوزان W_i بـالقيم المقـدرة لهـا $\widehat{W}_i$ حيث:

$$\widehat{W}_i = \frac{1}{\widehat{\sigma}_i}$$

وتتخذ مصفوفة التصحيح الصيغة التالية:

$$\hat{W}^{-1} = \begin{pmatrix} \frac{1}{\hat{\sigma}_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\hat{\sigma}_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\hat{\sigma}_3} & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \frac{1}{\hat{\sigma}_n} \end{pmatrix}$$

وعليه يتم الحصول على مقدرات المربعات الصغرى المرجحة ($\hat{\boldsymbol{\beta}}_{WLS}$) على النحو التالي:

$$\hat{\boldsymbol{\beta}}_{WLS} = (\mathbf{x}^T\hat{\mathbf{W}}^{-1}\mathbf{x})^{-1}\mathbf{x}^T\hat{\mathbf{W}}^{-1}\mathbf{y} \qquad (7\text{-}47)$$

وكذلك يتم الحصول على صيغة تباين معاملات الانحدار المقدرة باستخدام طريقة المربعات الصغرى المرجحة كما يلي:

$$\text{var}(\hat{\boldsymbol{\beta}}_{WLS}) = \hat{\sigma}^2(\mathbf{x}^T\mathbf{W}^{-1}\mathbf{x})^{-1} \qquad (7\text{-}48)$$

- من الطرق المستخدمة لعلاج مشكلة اختلاف التباين إجراء تحويلة للنموذج الأصلي حسب نمط اختلاف التباين. حيث تربط أحياناً بين تباين حد الخطأ (σ_i^2) وقيم المتغير المستقل علاقة منتظمة. ففي نموذج الانحدار الخطي البسيط مثلاً قد تأخذ العلاقة بين تباين حدود الخطأ (σ_i^2) والمتغير المستقل (X_i) إحدى الصيغ التالية:

$$\sigma_i^2 = \sigma^2 x_i \qquad i = 1,2,\dots,n \qquad (7\text{-}49)$$

$$\sigma_i^2 = \sigma^2 x_i^2 \qquad i = 1,2,\dots,n \qquad (7\text{-}50)$$

$$\sigma_i^2 = \sigma^2 \sqrt{x_i} \qquad x_i > 0 \qquad i = 1,2,\dots,n \qquad (7\text{-}51)$$

- الصيغة الأولى:

إذا بدا لنا من رسم شكل انتشار البواقي أو البواقي المعيارية مع المتغير المستقل أن تباين البواقي يتغير مع تغير قيم المتغير المستقل (X_i) بحيث إن قيمة $\hat{\sigma}_i^2$ تزداد بزيادة قيمة (X_i) كما في المعادلة (7.49)، فإنه يمكن استخدام الأوزان التالية لعلاج المشكلة:

$$\widehat{W}_i = \frac{1}{\sqrt{x_i}}$$

وفي هذه الحالة يأخذ نموذج الانحدار البسيط المرجح الشكل التالي:

$$\frac{y_i}{\sqrt{x_i}} = \frac{\beta_0}{\sqrt{x_i}} + \beta_1\sqrt{x_i} + \frac{\varepsilon_i}{\sqrt{x_i}}$$

أو

$$y_i^* = \beta_0 x_{1i}^* + \beta x_{2i}^* + \varepsilon_i^*$$

حيث إن $y_i^* = \frac{y_i}{\sqrt{x_i}}$, $x_{1i}^* = \frac{1}{\sqrt{x_i}}$, $x_{2i}^* = \sqrt{x_i}$, $\varepsilon_i^* = \frac{\varepsilon_i}{\sqrt{x_i}}$

ويلاحظ أن المتغير العشوائي (ε_i^*) يستوفي اشتراطات طريقة المربعات الصغرى ويمكن برهان ذلك كما يلي:

القيمة المتوقعة:

$$E(\varepsilon_i^*) = E\left(\frac{\varepsilon_i}{\sqrt{x_i}}\right) = 0$$

التباين:

$$\text{var}(\varepsilon_i^*) = E\left(\frac{\varepsilon_i^2}{x_i}\right) = \frac{\sigma^2 x_i}{x_i} = \sigma^2$$

وبعد ترجيح النموذج تستخدم طريقة المربعات الصغرى غير المرجحة على النموذج المرجح.

وفي حالة نموذج الانحدار الخطي المتعدد، إذا اتضح من رسم انتشار البواقي أو البواقي المعيارية مع المتغير المستقل أن العلاقة بين $\widehat{\sigma}_i^2$ والمتغير المستقل (X_j) مصدر اختلاف التباين تأخذ الصيغة التالية:

$$\sigma_i^2 = \sigma^2 x_{ji} \qquad i = 1,2,\ldots,n$$

ويتم ترجيح المتغيرات المستقلة والمتغير التابع باستخدام الأوزان التالية:

$$\widehat{W}_i = \frac{1}{\sqrt{x_{ji}}} \qquad i=1,2,\ldots,n$$

تتخذ مصفوفة التصحيح (الترجيح) في هذه الحالة الشكل التالي:

$$\widehat{\mathbf{W}}^{-1}=\begin{pmatrix} \frac{1}{\sqrt{x_{j1}}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{x_{j2}}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{x_{j3}}} & \dots & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{x_{jn}}} \end{pmatrix}$$

وعليه يمكن استخدام طريقة المربعات الصغرى المرجحة لتقدير معالم نموذج الانحدار الخطي المتعدد باستخدام الصيغة التالية:

$$\widehat{\boldsymbol{\beta}}_{WLS}=(\mathbf{x}^{T}\widehat{\mathbf{W}}^{-1}\mathbf{x})^{-1}\mathbf{x}^{T}\widehat{\mathbf{W}}^{-1}\mathbf{y}$$

ويمكن الحصول على نفس النتائج باستخدام النموذج المحوّل (المرجح) التالي:

$$\frac{y_i}{\sqrt{x_{ji}}}=\beta_0\frac{1}{\sqrt{x_{ji}}}+\beta_1\frac{x_{1i}}{\sqrt{x_{ji}}}+\dots+\beta_j+\dots+\beta_p\frac{x_{pi}}{\sqrt{x_{ji}}}+\frac{\varepsilon_i}{\sqrt{x_{ji}}}$$

وبعد ترجيح النموذج تستخدم طريقة المربعات الصغرى غير المرجحة على النموذج المرجح.

- الصيغة الثانية:

أما إذا أظهر رسم الانتشار أن تباين بواقي النموذج الأصلي المقدر($\hat{\sigma}_i^2$) يتغير مباشرة مع (x_i^2) كما في الصيغة (7.50) فإنه تستخدم الأوزان التالية:

$$\widehat{W}_i=\frac{1}{x_i}$$

ويصبح النموذج المحوّل على النحو التالي:

$$\frac{y_i}{x_i}=\frac{\beta_0}{x_i}+\beta_1+\frac{\varepsilon_i}{x_i}$$

أو

$$y_i^*=\beta_0 x_i^*+\beta_1+\varepsilon_i^*$$

حيث إن $y_i^*=\frac{y_i}{x_i},\quad x_i^*=\frac{1}{x_i},\quad \varepsilon_i^*=\frac{\varepsilon_i}{x_i}$

ويلاحظ أيضاً أن المتغير العشوائي (ε_i^*) يستوفي اشتراطات طريقة المربعات الصغرى، إذ نجد أن القيمة المتوقعة للمتغير العشوائي تساوي صفر، أي:

$$E(\varepsilon_i^*) = E\left(\frac{\varepsilon_i}{x_i}\right) = 0$$

وأن تباين المتغير العشوائي (ε_i^*) ثابت، أي:

$$\text{var}(\varepsilon_i^*) = E\left(\frac{\varepsilon_i}{x_i}\right)^2 = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2$$

وفي حالة نموذج الانحدار الخطي المتعدد، إذا تبين لنا من رسم انتشار البواقي أو البواقي المعيارية أن العلاقة بين ($\hat{\sigma}_i^2$) والمتغير المستقل (X_j) مصدر اختلاف التباين تأخذ الصيغة التالية:

$$\sigma_i^2 = \sigma^2 x_{ji}^2$$

تستخدم الأوزان التالية:

$$\hat{W}_i = \frac{1}{x_{ji}} \qquad i = 1,2,...,n$$

ويتم استبدال قيم W_i الواردة في مصفوفة التصحيح بالقيم المقدرة لها $\hat{W}_i$ ويتم إجراء حل النموذج للحصول على مقدرات المربعات الصغرى والإحصاءات الأخرى المرتبطة بها.

- **الصيغة الثالثة:**

إذا كانت العلاقة بين تباين البواقي والمتغير المستقل (X_i) تأخذ الصيغة التالية:

$$\hat{\sigma}_i^2 = \sigma^2 \sqrt{x_i}$$

فتستخدم الأوزان التالية:

$$\hat{W}_i = \frac{1}{x^{\frac{1}{4}}}$$

ويصبح النموذج الخطي البسيط المرجح كما يلي:

$$\frac{y_i}{x_i^{\frac{1}{4}}} = \frac{\beta_0}{x_i^{\frac{1}{4}}} + \frac{\beta_1 x_i}{x_i^{\frac{1}{4}}} + \frac{\varepsilon_i}{x_i^{\frac{1}{4}}}$$

أو

$$y_i^* = \beta_0 x_{1i}^* + \beta_1 x_{2i}^* + \varepsilon_i^*$$

حيث إن $y_i^* = \frac{y_i}{x_i^{\frac{1}{4}}}, \quad x_{1i}^* = \frac{1}{x_i^{\frac{1}{4}}}, \quad x_{2i}^* = \frac{x_i}{x_i^{\frac{1}{4}}}, \quad \varepsilon_i^* = \frac{\varepsilon_i}{x_i^{\frac{1}{4}}}$

ويلاحظ أيضاً أن المتغير العشوائي يستوفي فرضيات طريقة المربعات الصغرى، أي أن:

$$E(\varepsilon_i^*) = E\left(\frac{\varepsilon_i}{x_i^{\frac{1}{4}}}\right)^2 = 0$$

و

$$\text{var}(\varepsilon_i^*) = E\left(\frac{\varepsilon_i}{x_i^{\frac{1}{4}}}\right)^2 = \frac{E(\varepsilon_i)^2}{x_i^{\frac{1}{2}}} = \frac{\sigma^2\sqrt{x_i}}{\sqrt{x_i}} = \sigma^2$$

- إجراء تحويلة للمتغير التابع: من الحلول المقترحة لتثبيت تباين حد الخطأ إجراء تحويلة للمتغير التابع. فيما يلي بعض التحويلات المقترحة لتثبيت تباين حد الخطأ حسب كل حالة:

- التحويلة اللوغاريتمية ($y^* = \ln(y)$) وتستخدم في حالة أن تباين y يزداد بزيادة قيم y_i.
- تحويلة الجذر التربيعي ($y^* = \sqrt{y}$) وتستخدم في حالة أن تباين Y يزداد بزيادة الوسط الحسابي لـ Y. وتعد هذه التحويلة مناسبة في حالة أن المتغير التابع له توزيع بواسون (Poisson Distribution).
- التحويلة التربيعية ($y^* = y^2$)، وتستخدم في حالة أن تباين y يتناقص بزيادة الوسط الحسابي لـ y.

٧-٣-٦ مثال:

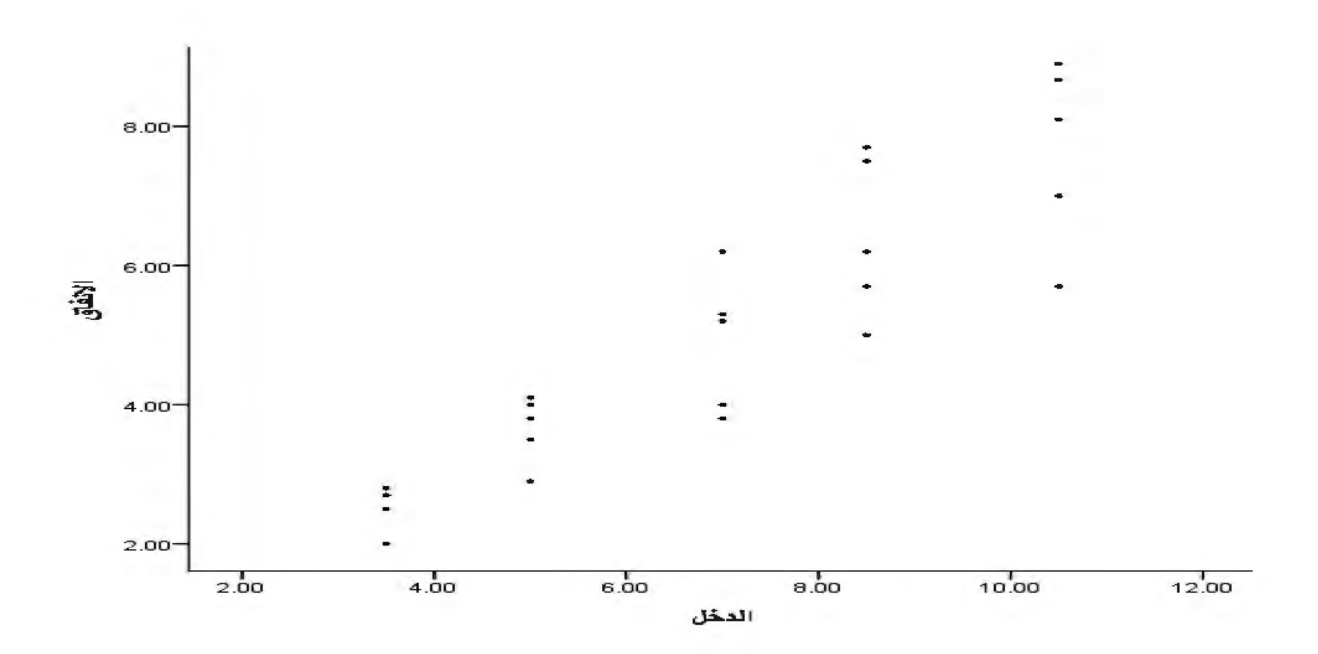
يوضح الجدول رقم (٧-٤) بيانات افتراضية عن الدخل والإنفاق الشهري بآلاف الريالات لعدد (٢٥) شخصاً. ويوضح رسم الانتشار (شكل رقم (٧-٥)) أن العلاقة بين الإنفاق والدخل الشهري علاقة خطية. وبإجراء انحدار الإنفاق الشهري (Y) على الدخل الشهري (X) نحصل على النموذج المقدر التالي:

$$\hat{y}_i = -0.08527 + 0.742619x_i \qquad R^2 = 0.824$$
$$(0.526) \qquad (0.072)$$
$$(0.873) \qquad (0.000)$$

حيث إن الأرقام الموجودة بين القوسين إلى أسفل هي قيم الأخطاء المعيارية لمقدرات معالم النموذج وقيم الاحتمال على التوالي. وتوضح نتائج النموذج أن هناك علاقة ذات دلالة إحصائية ($p\text{-value} << 0.01$) بين الإنفاق والدخل الشهري، وأن الدخل الشهري يفسر ٨٢,٤% من التغير في الإنفاق الشهري.

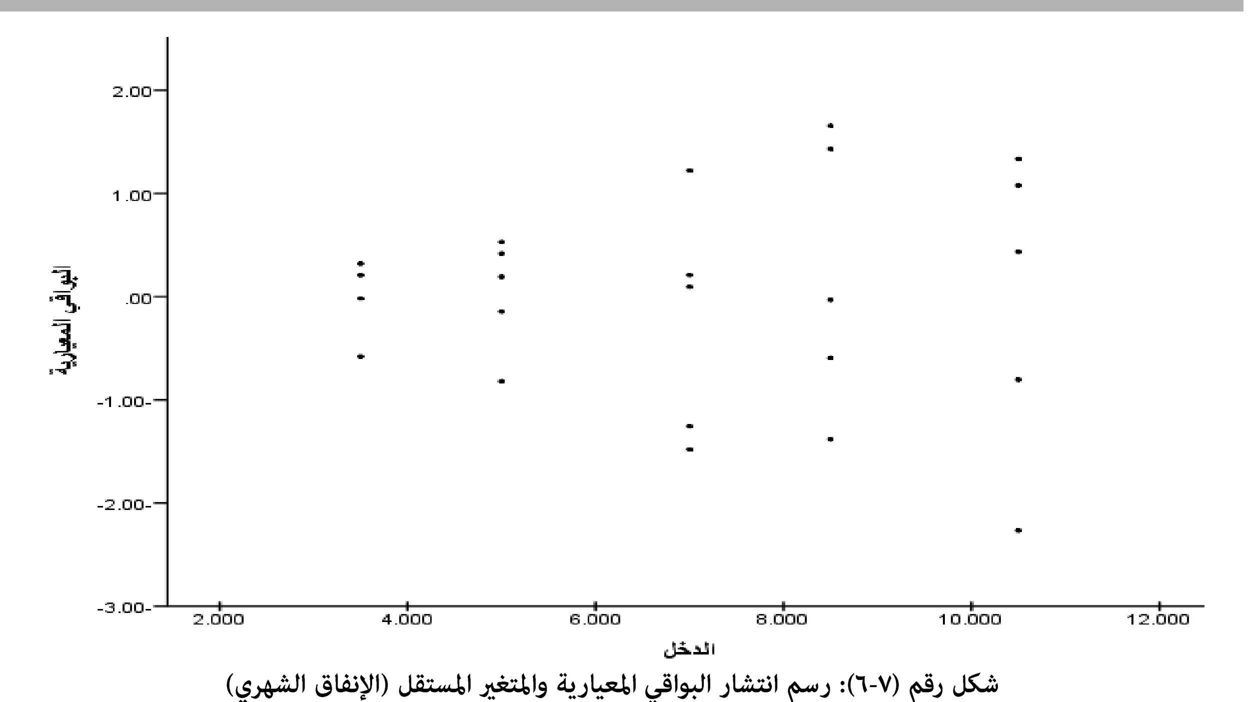
جدول رقم (٧-٤): بيانات افتراضية عن الدخل والإنفاق الشهري

الدخل الشهري (ألف ريال)	الإنفاق الشهري للفرد (ريال)				
3.5	2.00	2.70	2.70	2.50	2.80
5.0	4.10	3.50	3.80	4.00	2.90
7.0	5.20	5.30	4.00	3.80	6.20
8.5	5.70	6.20	7.50	7.70	5.00
10.5	8.10	5.70	8.90	8.67	7.00



شكل رقم (٧-٥): رسم انتشار الإنفاق الشهري مع الدخل الشهري

وبفحص النموذج المقدر يلاحظ أن تباين البواقي غير ثابت عند كل قيم المتغير المستقل كما يظهر ذلك الشكل رقم (٧-٦).



شكل رقم (٧-٦): رسم انتشار البواقي المعيارية والمتغير المستقل (الإنفاق الشهري)

وللتحقق أكثر من وجود عدم ثبات التباين تم إجراء الاختبارات التالية:

- **اختبار جولدفيلد-كواندت:**

لإجراء هذا الاختبار، تم أولاً ترتيب المشاهدات حسب قيم المتغير المستقل (الدخل الشهري) تصاعدياً ومن ثم تم استبعاد الخمس مشاهدات الوسطى (مجموعة مشاهدات الدخل المساوي لـ ٧ ألف ريال) ومن بعد ذلك تم بناء نموذجي انحدار، الأول للقيم الصغيرة لـ X_i والثاني للقيم الكبيرة. وفيما يلي نتائج النموذجين:

نتائج النموذج الأول:

$$\hat{y}_i = -0.07333 + 0.746667x_i, \quad R^2 = 0.70, \quad RSS_1 = 1.344$$

نتائج النموذج الثاني:

$$\hat{y}_i = 1.0905 + 0.627x_i, \quad R^2 = 0.24, \quad RSS_2 = 12.41552$$

والآن يتم إيجاد القيمة المحسوبة لإحصائية F كما يلي:

$$F_1 = \frac{RSS_2/V_2}{RSS_1/V_1} = \frac{12.41552}{1.344} = 9.238$$

وبما أن قيمة F المحسوبة أكبر من قيمة F الجدولية ($F_{0.01,8,8}=6.03$) عند مستوى معنوية ١% فإننا نرفض فرض العدم، أي أنه يوجد اختلاف في تباين حد الخطأ.

- **اختبار بارك:**

لإجراء هذا الاختبار تم تحويل المتغير المستقل إلى لوغاريتم $\ln(x)$ وتم حساب لوغاريتم مربع بواقي النموذج الأصلي $\ln(e_i^2)$ كما موضح بالجدول رقم (٧-٥)، ومن ثم تم بناء نموذج انحدار $\ln(e_i^2)$ على $\ln(x)$ وتم الحصول على النموذج المقدر التالي:

$$\widehat{\ln}(e_i^2) = -7.4995 + 3.0444\ln(x_i), \qquad R^2 = 0.232$$
$$(2.1964) \qquad (1.1559)$$
$$(0.002) \qquad (0.015)$$

حيث إن الأرقام الموجودة بين القوسين إلى أسفل هي قيم الأخطاء المعيارية لمقدرات معالم النموذج وقيم الاحتمال على التوالي. وتوضح نتائج هذا النموذج أن هناك علاقة ذات دلالة إحصائية بين $\ln(e_i^2)$ و $\ln(x)$ (P-value =0.015) مما يدعم وجود اختلاف في تباين حد الخطأ.

جدول رقم (٧-٥): بيانات المتغير المستقل، البواقي، لوغاريتم مربع البواقي، ولوغاريتم المتغير المستقل

رقم المشاهدة	x_i	البواقي (e_i)	$\ln(e_i^2)$	$\ln(x)$
1	3.5	-0.513895765	-1.33146965	1.252762968
2	3.5	0.186104235	-3.362896723	1.252762968
3	3.5	0.186104235	-3.362896723	1.252762968
4	3.5	-0.013895765	-8.552342255	1.252762968
5	3.5	0.286104235	-2.502798156	1.252762968
6	5.0	0.472175896	-1.500807405	1.609437912
7	5.0	-0.127824104	-4.114200292	1.609437912
8	5.0	0.172175896	-3.51847735	1.609437912
9	5.0	0.372175896	-1.976777397	1.609437912
10	5.0	-0.727824104	-0.63539175	1.609437912
11	7.0	0.086938111	-4.885117568	1.945910149
12	7.0	0.186938111	-3.353955351	1.945910149
13	7.0	-1.113061889	0.214229353	1.945910149
14	7.0	-1.313061889	0.54472346	1.945910149
15	7.0	1.086938111	0.166729341	1.945910149
16	8.5	-0.526990228	-1.281146547	2.140066163
17	8.5	-0.026990228	-7.224560808	2.140066163

رقم المشاهدة	x_i	البواقي (e_i)	$\ln(e_i^2)$	$\ln(x)$
18	8.5	1.273009772	0.482767992	2.140066163
19	8.5	1.473009772	0.774615543	2.140066163
20	8.5	-1.226990228	0.409128403	2.140066163
21	10.5	0.387771987	-1.894675549	2.351375257
22	10.5	-2.012228013	1.398485145	2.351375257
23	10.5	1.187771987	0.344158545	2.351375257
24	10.5	0.957771987	-0.086291078	2.351375257
25	10.5	-0.712228013	-0.678714352	2.351375257

- **اختبار جليجسر (Glejeser Test):**

لإجراء هذا الاختبار تم حساب القيم المطلقة لبواقي النموذج الأصلي (جدول رقم (٧-٦))، ومن ثم تم إجراء انحدار البواقي المطلقة ($|e_i|$) على المتغير المستقل (X_i). حيث تم التوصل إلى النموذج التالي:

$$|\hat{e}_i| = -0.187073 + 0.123518x_i, \qquad R^2 = 0.33$$

$$(0.268761) \quad (0.036659)$$

$$(0.493) \quad (0.003)$$

حيث إن الأرقام الموجودة بين القوسين إلى أسفل هي قيم الأخطاء المعيارية لمقدرات معالم النموذج وقيم الاحتمال على التوالي. وتوضح نتائج هذا النموذج أن هناك علاقة ذات دلالة إحصائية بين المتغير المستقل (X) والقيم المطلقة للبواقي (P-value <0.01) مما يجعلنا نرفض فرض العدم ونقبل الفرض البديل القائل بأن تباين حد الخطأ غير ثابت.

جدول رقم (٧-٦): المتغير المستقل، البواقي، والبواقي المطلقة.

رقم المشاهدة	x_i	البواقي (e_i)	القيمة المطلقة للبواقي
1	3.5	-0.5139	0.513895765
2	3.5	0.186104	0.186104235
3	3.5	0.186104	0.186104235
4	3.5	-0.0139	0.013895765
5	3.5	0.286104	0.286104235
6	5.0	0.472176	0.472175896
7	5.0	-0.12782	0.127824104
8	5.0	0.172176	0.172175896
9	5.0	0.372176	0.372175896
10	5.0	-0.72782	0.727824104
11	7.0	0.086938	0.086938111
12	7.0	0.186938	0.186938111
13	7.0	-1.11306	1.113061889
14	7.0	-1.31306	1.313061889

القيمة المطلقة للبواقي	البواقي (e_i)	x_i	رقم المشاهدة
1.086938111	1.086938	7.0	15
0.526990228	-0.52699	8.5	16
0.026990228	-0.02699	8.5	17
1.273009772	1.27301	8.5	18
1.473009772	1.47301	8.5	19
1.226990228	-1.22699	8.5	20
0.387771987	0.387772	10.5	21
2.012228013	-2.01223	10.5	22
1.187771987	1.187772	10.5	23
0.957771987	0.957772	10.5	24
0.712228013	-0.71223	10.5	25

- **اختبار بروش-باقان** (Breusch & Pagan Test):

لإجراء هذا الاختبار تمت الخطوات التالية:

-تم إجراء نموذج الانحدار الأصلي وتم الحصول على البواقي (انظر الجدول رقم (٧-٧)) ومن ثم تم حساب التباين المتحيز كما يلي:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{25} e_i^2}{n} = \frac{18.145865}{25} = 0.726$$

- ومن ثم تم حساب قيم السلسلة التالية (انظر الجدول):

$$g_i = \frac{e_i^2}{\hat{\sigma}^2} \qquad i = 1,2,\ldots,25$$

- تم إجراء نموذج انحدار g_i على المتغير x والحصول على النتائج التالية:

$$\hat{g}_i = -0.94864 + 0.28241 x_i, \quad R^2 = 0.2887, \quad ESS = 12.24258$$
$$(0.677567) \qquad (0.092419)$$

حيث إن الأرقام الموجودة بين القوسين إلى أسفل هي قيم الأخطاء المعيارية لمقدرات معالم النموذج.

- ومن نتائج النموذج نجد أن مجموع مربعات الانحدار يساوي (١٢,٢٤٢٥٨).

- وحيث إن نصف مجموع مربعات الانحدار أكبر من قيمة توزيع مربع كاي بدرجة حرية واحدة ومستوى معنوية ٥%، فإننا نرفض فرض العدم ونقبل الفرض البديل بأن تباين حد الخطأ غير ثابت عند كل قيم المتغير المستقل.

$$Q=\frac{ESS}{2}=\frac{12.24258}{2}=6.1212991>\chi^2_{1,0.05}=3.8414$$

ملحوظة:

يتم حساب اختبار بروش - باقان في برنامج EVIEWS باستخدام الصيغة التالية:

$$Q=n\times R^2 \sim \chi^2_m \qquad (7\text{-}52)$$

حيث إن n عدد المشاهدات و R^2 معامل التحديد لنموذج انحدار مربعات بواقي النموذج الأصلي (e_i^2) على المتغيرات المستقلة التي يعتقد أنها سبب تباين الاختلاف ($Z_2\ Z_3 \ldots Z_m$). ومن ثم يتم إجراء الاختبار الإحصائي بمقارنة قيمة Q بالقيمة الجدولية لتوزيع مربع كاي عند درجة حرية (m) -عدد المتغيرات المستقلة-ومستوى معنوية محدد. فمثلاً من المثال نجد أن قيمة Q كما يلي:

$$Q=n\times R^2 = 25\times 0.288755=7.2189$$

- **اختبار وايت** (White's Test):

لإجراء هذا الاختبار تمت الخطوات التالية:

- تم إجراء نموذج الانحدار الأصلي وتم الحصول على البواقي ومربعات البواقي (انظر الجدول رقم (٨-٦)).

- تم إجراء نموذج انحدار e_i^2 على المتغير x و x^2 والحصول على النتائج التالية:

$$\hat{e}_i^2=-0.84767+0.257204X_i - 0.00374X_i^2, \qquad R^2=0.289169$$

$$\begin{matrix} (1.49289) & (0.4664) & (0.0331) \\ (0.576) & (0.587) & (0.911) \end{matrix}$$

حيث إن الأرقام الموجودة بين القوسين إلى أسفل هي قيم الأخطاء المعيارية لمقدرات معالم النموذج وقيم الاحتمال على التوالي.

- ومن نتائج النموذج نجد معامل التحديد يساوي (٠,٢٨٩١٦٩).

- وحيث إن نصف قيمة معامل التحديد أكبر من قيمة توزيع مربع كاي بدرجة حرية واحدة ومستوى معنوية ٥%، فإننا نرفض فرض العدم ونقبل الفرض البديل بأن تباين حد الخطأ غير ثابت عند كل قيم المتغير المستقل.

$$Q=n\times R^2=25\times 0.289169=7.229216>\chi^2_{2,0.05}=5.99$$

جدول (٧-٧): المتغير المستقل والبواقي ومربع البواقي وقيمة g_i.

g_i	e_i^2	e_i	x^2	x	رقم المشاهدة
0.363786	0.264089	-0.5139	12.25	3.5	1
0.04771	0.034635	0.186104	12.25	3.5	2
0.04771	0.034635	0.186104	12.25	3.5	3
0.000266	0.000193	-0.0139	12.25	3.5	4
0.112757	0.081856	0.286104	12.25	3.5	5
0.307117	0.22295	0.472176	25.00	5	6
0.022507	0.016339	-0.12782	25.00	5	7
0.040836	0.029645	0.172176	25.00	5	8
0.190806	0.138515	0.372176	25.00	5	9
0.729707	0.529728	-0.72782	25.00	5	10
0.010412	0.007558	0.086938	49.00	7	11
0.048138	0.034946	0.186938	49.00	7	12
1.70661	1.238907	-1.11306	49.00	7	13
2.375013	1.724132	-1.31306	49.00	7	14
1.627441	1.181434	1.086938	49.00	7	15
0.382561	0.277719	-0.52699	72.25	8.5	16
0.001003	0.000728	-0.02699	72.25	8.5	17
2.232334	1.620554	1.27301	72.25	8.5	18
2.988869	2.169758	1.47301	72.25	8.5	19
2.073852	1.505505	-1.22699	72.25	8.5	20
0.207133	0.150367	0.387772	110.25	10.5	21
5.577634	4.049062	-2.01223	110.25	10.5	22
1.943398	1.410802	1.187772	110.25	10.5	23
1.26363	0.917327	0.957772	110.25	10.5	24
0.698769	0.507269	-0.71223	110.25	10.5	25

معالجة مشكلة عدم ثبات تباين حد الخطأ:

أوضحت الاختبارات التي تم تطبيقها أن نموذج الإنفاق يعاني من مشكلة عدم ثبات تباين حد الخطأ العشوائي. ولعلاج هذه المشكلة تم حساب التباين والانحراف المعياري لبواقي نموذج الانحدار الأصلي عند كل قيمة من قيم المتغير المستقل (X) ومن ثم تم حساب بعض الأوزان المُرشحة لتثبيت التباين (جدول رقم (٧-٨)).

جدول رقم (٧-٨): قيم المتغير المستقل، تباين بواقي النموذج الأصلي المناظرة لهذه القيم وبعض الأوزان المرشحة لتثبيت تباين حد الخطأ.

المجموعة	1	2	3	4	5
مستويات المتغير المستقل	3.5	5.0	7.0	8.5	10.5
تباين البواقي S_j^2	0.10300	0.23300	0.99000	1.34700	1.75688
الانحراف المعياري للبواقي S_j	0.32094	0.48270	0.99499	1.16060	1.32547
$1/S_j^2$	9.71	4.29	1.01	0.74	0.57
$1/S_j$	3.12	2.07	1.01	0.86	0.75
S_j^2/x_j	0.03	0.05	0.14	0.16	0.17
S_j^2/x_j^2	0.00840	0.00930	0.02020	0.01860	0.01590
$S_j^2/\sqrt{x_j}$	0.05506	0.10420	0.37418	0.46202	0.54218

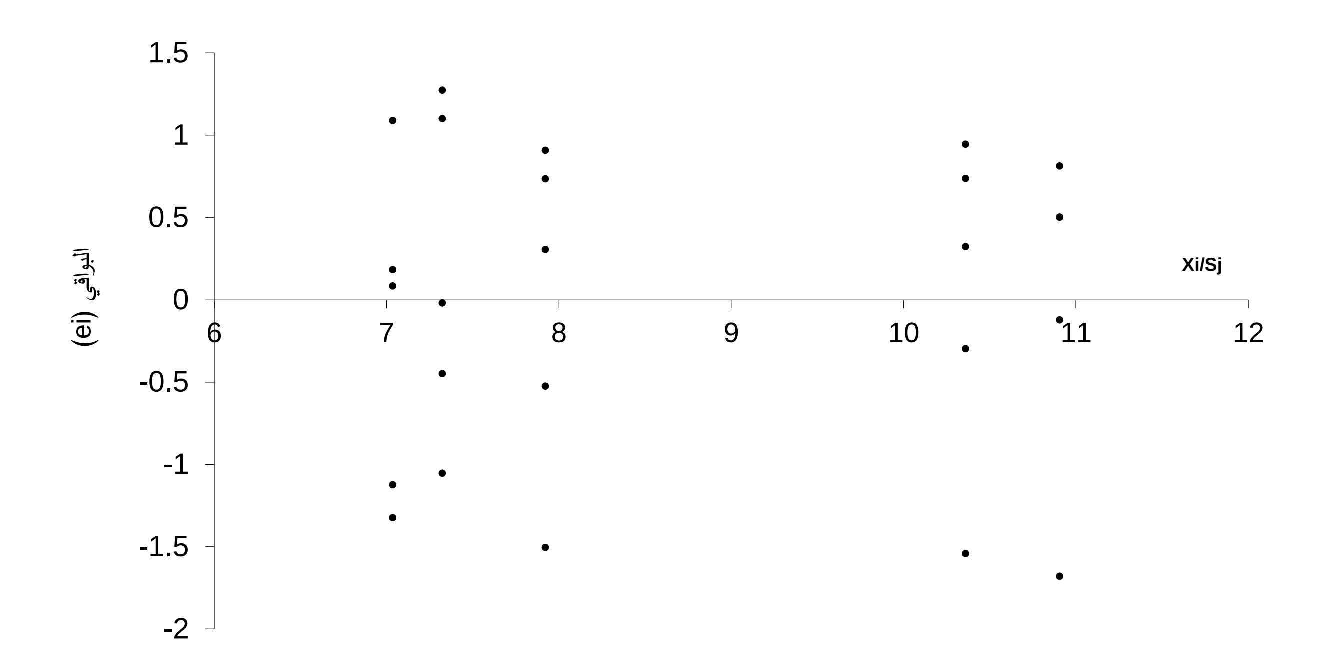
وبملاحظة الجدول وجد أن أفضل مرجح (وزن) يمكن استخدامه لتثبيت التباين هو معكوس المتغير الانحراف المعياري ($\frac{1}{S_j}$). وباستخدام هذا المرجح يمكن الحصول على نموذج الانحدار تم التوصل على النموذج المقدر التالي:

$$\hat{y}_i = -0.05374 + 0.7386x, \qquad R^2 = 0.867$$
$$(0.3647) \quad (0.0604)$$
$$(0.884) \quad (0.000)$$

حيث إن الأرقام الموجودة بين القوسين إلى أسفل هي قيم الأخطاء المعيارية والاحتمال لمقدرات معالم النموذج. ويلاحظ أن هناك اختلافاً طفيفاً بين قيمتي المعامل ($\hat{\beta}_1$) التي تم الحصول عليها عن طريق المربعات الصغرى الاعتيادية وتلك التي تم الحصول عليها باستخدام طريقة المربعات الصغرى المرجحة (الجدول رقم(٧-٩)). كما يلاحظ أن قيم الخطأ المعياري لمعاملي النموذج قد انخفضت في نموذج الانحدار المرجح. كما ارتفع معامل التحديد في النموذج المرجح من (٠,٨٢٤) إلى (٠,٨٦٧). وتبين نتائج اختبارات الكشف عن عدم ثبات التباين أن فرض عدم ثبات التباين مستوف في النموذج المرجح كما يوضح ذلك اختبار وايت $(p\text{-value} > 0.05)$ والشكل رقم (٧-٧).

جدول رقم (٧-٩): نتائج نموذجي الانحدار العادي والمرجح

المعامل	نموذج الانحدار العادي (غير المرجح)			نموذج الانحدار المرجح		
	قيمة المعامل	الخطأ المعياري	قيمة الاحتمال	قيمة المعامل	الخطأ المعياري	قيمة الاحتمال
$\widehat{\beta}_0$	-0.08527	0.525646	0.873	-0.05374	0.36472	0.884
$\widehat{\beta}_1$	0.742619	0.071698	0.000	0.7386	0.06041	0.000
R^2	0.824			0.867		
اختبار وايت	$Q = 7.23 > \chi^2_{2,0.05} = 5.99$			$Q = 4.33 < \chi^2_{2,0.05} = 5.99$		



شكل رقم (٧-٧): رسم انتشار بواقي النموذج المرجح مع المتغير المستقل

٧-٤ الارتباط الذاتي:

٧-٤-١ مقدمة:

تتضمن الاشتراطات الأساسية لطريقة المربعات الصغرى فرض عدم وجود ارتباط ذاتي أو تسلسلي بين حدود الخطأ، أي:

$$E(\varepsilon_i\varepsilon_j) = \sigma^2 \qquad i = j$$
$$= 0 \qquad i \neq j$$

وفي حالة عدم استيفاء هذا الاشتراط نواجه بمشكلة تعرف بالارتباط الذاتي (Autocorrelation). ويشير الارتباط الذاتي بوجه عام إلى وجود ارتباط بين قيم المشاهدات المتسلسلة لنفس المتغير خلال فترة زمنية محددة. وفي تحليل الانحدار يشير الارتباط الذاتي إلى وجود ارتباط بين قيم حدود الخطأ المتتالية، أي أن:

$$E(\varepsilon_i\varepsilon_j) \neq 0, \qquad i \neq j \quad i,j = 1,2,\ldots,N \qquad (7\text{-}51)$$

أو باستخدام رموز المصفوفات، نجد أن:

$$E(\varepsilon\varepsilon^T) = \begin{pmatrix} \sigma^2 & E(\varepsilon_1\varepsilon_2) & E(\varepsilon_1\varepsilon_3) & \ldots & E(\varepsilon_1\varepsilon_N) \\ E(\varepsilon_2\varepsilon_1) & \sigma^2 & E(\varepsilon_2\varepsilon_3) & \ldots & E(\varepsilon_2\varepsilon_N) \\ E(\varepsilon_3\varepsilon_1) & E(\varepsilon_3\varepsilon_2) & \sigma^2 & \ldots & E(\varepsilon_3\varepsilon_N) \\ . & . & . & \ldots & . \\ E(\varepsilon_N\varepsilon_1) & E(\varepsilon_N\varepsilon_2) & E(\varepsilon_N\varepsilon_3) & \ldots. & \sigma^2 \end{pmatrix} \qquad (7\text{-}53)$$

حيث إن:

$$E(\varepsilon_i\varepsilon_j) \neq 0 \qquad i \neq j$$
$$= \sigma^2 \qquad i = j$$

وتبين مصفوفة تباين وتغاير حد الخطأ أن قيم معامل الارتباط أو معامل التغاير بين حدود الخطأ غير مساوية للصفر. ويأخذ الارتباط الذاتي بين حدود الخطأ أشكالاً مختلفة. فقد يكون الارتباط الذاتي من الرتبة الأولى (First-Order Autoregressive Process (AR(1))) ، وفي هذه الحالة تكون كل قيمة من قيم حدود الخطأ مرتبطة بالقيمة التي تسبقها فقط حيث تأخذ العلاقة الصيغة التالية:

$$\varepsilon_t = \rho\varepsilon_{t-1} + a_t, \qquad |\rho| < 1 \qquad (7\text{-}54)$$

حيث إن:

ε_t = حد الخطأ في نموذج الانحدار الخطي المتعدد ($y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_{pt} x_{pt} + \varepsilon_t$)

ρ = معامل الارتباط الذاتي.

a_t = متغير عشوائي يتبع التوزيع الطبيعي بوسط حسابي يساوي الصفر وتباين ثابت σ_a^2 وأنه مستقل عن ε_t، كما يمكن أن تكون العلاقة بين حدود الخطأ العشوائي من الرتبة الثانية (Second-Order Autoregressive Process (AR(2))، وفي هذه الحالة تكون كل قيمة من قيم حدود الخطأ العشوائي مرتبطة بالقيمتين السابقتين لها، أي أن:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + a_t \qquad (7\text{-}55)$$

وهكذا قد تأخذ العلاقة بين حدود الخطأ العشوائي رتبة أعلى. إلا أنه في الواقع العملي يفترض نمط الارتباط الذاتي من الرتبة الأولى (Studenmud (2010), p.311 و Fox (1997) p.373).

٧-٤-٢ أسباب وجود الارتباط الذاتي:

يرجع وجود الارتباط الذاتي بين حدود الخطأ العشوائي إلى عدة أسباب نذكر منها ما يلي:

- عدم إدراج متغير/متغيرات مفسرة مؤثرة في المتغير التابع في النموذج. فإذا كانت قيم هذا المتغير/المتغيرات مرتبطة ذاتياً فإن هذا ينعكس على المتغير العشوائي، لأن المتغير العشوائي في هذه الحالة يحتوي تأثير التغيرات العشوائية زائداً أثر المتغير/المتغيرات المحذوفة. فمثلاً إذا كان النموذج الحقيقي يضم ثلاثة متغيرات كما يلي:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t$$

وتم بناء نموذج يضم المتغيرين X_1 و X_2، أي:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$$

فإذا كانت قيم المتغير المحذوف (X_3) مرتبطة ذاتياً عبر فترات متتالية، فإنه يتوقع أن تكون حدود الخطأ في النموذج الثاني مرتبطة ذاتياً.

- ومن أسباب الارتباط الذاتي وجود خطأ في توصيف النموذج المقدر من حيث الشكل الجبري للدالة، فقد تكون العلاقة بين المتغير التابع والمتغيرات المفسرة علاقة غير خطية في حين يتم بناء النموذج على أساس أن العلاقة خطية. فعلى سبيل المثال إذا كان النموذج الحقيقي هو:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$$

وتم توفيق النموذج التالي:

$$y_t = \beta_0 + \beta_1 x_t + v_t$$

فإنه يمكن أن نواجه بمشكلة ارتباط ذاتي بين حدود المتغير العشوائي (V_t).

- عدم إدراج المتغير التابع كمتغير متباطئ (Lagged variable) في الحالات التي تتطلب ذلك. فمثلاً من المعلـوم أن استهلاك الآيس الكريم يتأثر بمتغيرات عديدة (درجات الحرارة، السعر، .. إلخ) من ضمنها مسـتوى الاسـتهلاك في الفترة السابقة. وفي هذه الحالة توجد ضرورة لإدخال متغير الاستهلاك في فترة سـابقة كمتغـير مفسرـ للاسـتهلاك كما في المثال التالي:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \varepsilon_t$$

حيث إن:

y_t = استهلاك الآيس كريم في الفترة t.

y_{t-1} = استهلاك الآيس كريم في الفترة السابقة (t-1).

x_t = متوسط درجات الحرارة.

فإذا لم يتم إدراج المتغير المتباطئ y_{t-1} كمتغير مفسر، فإنه يتوقع أن تكون حدود الخطأ مترابطة. ويرجـع السـبب في إدراج الاستهلاك في فترة سابقة كمتغير مفسر إلى أن المستهلكين لا يميلون إلى تغيير سلوكهم الاستهلاكي باستمرار.

- طبيعية بيانات السلاسل الزمنية: من المعلوم أن بعض متغيرات السلاسل الزمنية كالناتج المحلي الإجمالي، الأرقـام القياسية للأسـعار، الإنتـاج، مسـتويات البطالـة وغيرهـا غالبـاً مـا تتغـير سـوياً في فـترات الرخـاء وفـترات الركـود الاقتصادي. ولذلك من المتوقع أن نواجه بمشكلة الارتباط الذاتي في حالة بناء نموذج انحـدار يتضـمن مثـل هـذه المتغيرات.

- ومن أسباب الارتباط الذاتي وجود خطأ منتظم في قياس بعض متغيرات النموذج.

٧-٤-٣ خصائص حدود الخطأ في حالة الارتباط الذاتي من الرتبة الأولى ((AR(1)):

الوسط الحسابي:

كما أوضحنا من قبل أن العلاقة بين حدود الخطأ في حالة الارتباط الذاتي من الرتبة الأولى تأخذ الصيغة التالية:

$$\varepsilon_t = \rho\varepsilon_{t-1} + a_t$$

ويمكن كتابة هذه المعادلة بدلالة حدود الخطأ في الفترة السابقة أيضاً على النحو التالي:

$$\varepsilon_{t-1} = \rho\varepsilon_{t-2} + a_{t-1}$$

وبوضع هذه المعادلة في سابقتها نحصل على:

$$\varepsilon_t = \rho(\rho\varepsilon_{t-2} + a_{t-1}) + a_t = \rho^2\varepsilon_{t-2} + \rho a_{t-1} + a_t$$

وبما أن

$$\varepsilon_{t-2} = \rho\varepsilon_{t-3} + a_{t-2}$$

فإن

$$\varepsilon_t = \rho^2(\rho\varepsilon_{t-3} + a_{t-2}) + \rho a_{t-1} + a_t$$
$$= \rho^3\varepsilon_{t-3} + \rho^2 a_{t-2} + \rho a_{t-1} + a_t$$

وبالاستمرار في هذه العملية يمكن التعبير عن الارتباط الذاتي من الرتبة الأولى كما يلي:

$$\varepsilon_t = a_t + \rho a_{t-1} + \rho^2 a_{t-2} + \rho^3 a_{t-3} + ...$$
$$= \sum_{u=0}^{\infty} \rho^u a_{t-u} \qquad (7\text{-}56)$$

أي أن الارتباط الذاتي من الرتبة الأولى AR(1) عبارة عن مجموع التغير العشوائي الحالي وعدد لا نهائي مـن الحـدود التي تتضمن تغيرات عشوائية سابقة. وبأخذ القيمة المتوقعة للمعادلة (7.56) نحصل على:

$$E(\varepsilon_t) = E(\sum_{u=0}^{\infty} \rho^u a_{t-u})$$
$$= \sum_{u=0}^{\infty} \rho^u E(a_{t-u}) = 0$$

مما يدل على أن شرط التوقع الصفري لحد الخطأ ما يزال مستوفياً تحت ظل الارتباط الذاتي.

التباين:

يمكن إيجاد التباين على النحو التالي:

$$\text{var}(\varepsilon_t) = E(\varepsilon_t^2)$$
$$= E[(\sum_{u=0}^{\infty} \rho^u a_{t-u})^2]$$
$$= E[\sum_{u=0}^{\infty} \rho^{2u} a_{t-u}^2 + 2\sum_{u \neq r} \rho^u \rho^r a_{t-u} a_{t-r}]$$
$$= \sum_{u=0}^{\infty} \rho^{2u} E(a_{t-u}^2) + 2\sum_{u \neq r} \rho^u \rho^r E(a_{t-u} a_{t-r})$$
$$= \sigma_a^2 \sum_{u=0}^{\infty} \rho^{2u} + 0$$

وبما أن

$$\sum_{u=0}^{\infty}\rho^{2u}=\frac{1}{1-\rho^2}$$

لأن هذا الحد عبارة عن متوالية هندسية لا نهائية، فإن

$$\text{var}(\varepsilon_t)=\frac{\sigma_a^2}{1-\rho^2}$$

ويلاحظ أن الطرف الأيمن للمعادلة لا يحتوي على الدليل t مما يدل على أن السلسلة ε_t لها تباين ثابت هو:

$$\text{var}(\varepsilon_t)=\frac{\sigma_a^2}{1-\rho^2}=\sigma^2 \qquad (7\text{-}57)$$

التغاير:

أما التغاير بين القيم المتتالية لحدود الخطأ فيمكن حسابه كما يلي:

$$\begin{aligned}
\text{cov}\left(\varepsilon_t\varepsilon_{t-1}\right) &= E\left(\varepsilon_t\varepsilon_{t-1}\right)\\
&= E\left\{\left(a_t+\rho a_{t-1}+\rho^2 a_{t-2}+\rho^3 a_{t-3}+\ldots\right)\left(a_{t-1}+\rho a_{t-2}+\rho^2 a_{t-3}+\rho^3 a_{t-4}+\ldots\right)\right\}\\
&= E\left(a_t a_{t-1}+\rho a_t a_{t-2}+\rho^2 a_t a_{t-3}+\rho^3 a_t a_{t-4}+\ldots\right)+E\left(\rho a_{t-1}^2+\rho^2 a_{t-1}a_{t-2}+\rho^3 a_{t-1}a_{t-3}+\ldots\right)+\\
&\quad E\left(\rho^2 a_{t-2}a_{t-1}+\rho^3 a_{t-2}^2+\rho^4 a_{t-2}a_{t-3}+\rho^5 a_{t-2}a_{t-4}+\ldots\right)+\\
&\quad E\left(\rho^3 a_{t-3}a_{t-1}+\rho^4 a_{t-3}a_{t-2}+\rho^5 a_{t-3}^2+\rho^6 a_{t-3}a_{t-4}+\ldots\right)+\ldots
\end{aligned}$$

وبما أن:

$$E(a_t a_{t-u})=0 \qquad \text{for } u\neq 0$$

و

$$E(a_t)=0$$

فإن:

$$\begin{aligned}
\text{cov}\left(\varepsilon_t\varepsilon_{t-1}\right) &= 0+\rho\sigma_a^2+\rho^3\sigma_a^2+\rho^5\sigma_a^2+\ldots\\
&= \rho\sigma_a^2\left(1+\rho^2+\rho^4+\ldots\right)
\end{aligned}$$

وبما أن الحد $1+\rho^2+\rho^4+...$ عبارة عن متوالية هندسية لا نهائية فإن مجموعه يساوي:

$$1+\rho^2+\rho^4+...=\frac{1}{1-\rho^2}$$

وبالتالي فإن:

$$\text{cov}(\varepsilon_t\varepsilon_{t-1})=\frac{\rho\sigma_a^2}{1-\rho^2}=\rho\sigma^2$$

أي أن قيمة التغاير بين حد الخطأ في الفترة t والفترة السابقة (t-1) غير مساوية للصفر.

وباتباع نفس الطريقة يمكننا إيجاد تغاير $(\varepsilon_t\varepsilon_{t-1})$ على النحو التالي:

$$\begin{aligned}\text{cov}(\varepsilon_t\varepsilon_{t-2}) &= E\{(a_t+\rho a_{t-1}+\rho^2 a_{t-2}+...)(a_{t-2}+\rho a_{t-3}+\rho^2 a_{t-4}+...)\}\\ &=\rho^2\sigma_a^2+\rho^4\sigma_a^2+\rho^6\sigma_a^2+...\\ &=\rho^2\,\frac{\sigma_a^2}{1-\rho^2}=\rho^2\sigma^2\end{aligned}$$

وهكذا بصورة عامة يمكننا إيجاد تغاير $(\varepsilon_t\varepsilon_{t-s})$ كما يلي:

$$\text{cov}(\varepsilon_t\varepsilon_{t-s})=E(\varepsilon_t\varepsilon_{t-s})=\frac{\rho^s\sigma_a^2}{1-\rho^2}=\rho^s\sigma^2,\quad \text{for } s=1,2,...,N\text{-}1 \qquad (7\text{-}58)$$

والآن يمكن كتابة مصفوفة تباين-تغاير حد الخطأ كما يلي:

$$E\left(\varepsilon\varepsilon^T\right)=\begin{pmatrix}\sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & ... & \rho^{N-1}\sigma^2\\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & ... & \rho^{N-2}\sigma^2\\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & ... & \rho^{N-3}\sigma^2\\ . & . & . & ... & .\\ \rho^{N-1}\sigma^2 & \rho^{N-2}\sigma^2 & \rho^{N-3}\sigma^2 & ... & \sigma^2\end{pmatrix} \qquad (7\text{-}59)$$

أو

$$E(\varepsilon\varepsilon^T)=\sigma^2\Omega \qquad (7\text{-}60)$$

حيث إن:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ . & . & . & \dots & . \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix}$$

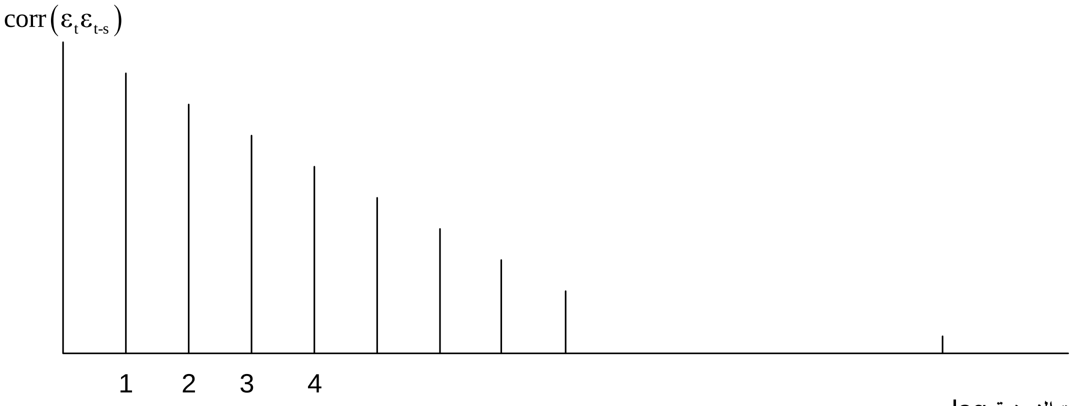
ويتضح من المصفوفة أعلاه أن التغاير بين حدود الخطأ لم يعد مساوياً للصفر في ظل الارتباط الذاتي.

معامل الارتباط بين حدود الخطأ:

يتم حساب معامل الارتباط بين ε_t و ε_{t-s} كما يلي:

$$\begin{aligned} corr(\varepsilon_t \varepsilon_{t-s}) &= \frac{cov(\varepsilon_t \varepsilon_{t-s})}{\sqrt{var(\varepsilon_t)}.\sqrt{var(\varepsilon_{t-s})}} \\ &= \frac{\rho^s \sigma^2}{\sigma.\sigma} = \rho^s \end{aligned} \quad (7-61)$$

وتبين هذه المعادلة أن قيم معاملات الارتباط الذاتي تقترب من الصفر كلما زاد طول الفجوة الزمنية (s) (الشكل ٧-٨).



شكل رقم (٧-٨): الارتباط الذاتي من الرتبة الأولى لحدود الخطأ العشوائي AR(1).

٧-٤-٤ النتائج المترتبة على وجود الارتباط الذاتي:

تظل مقدرات المربعات الصغرى غير متحيزة في ظل وجود الارتباط الذاتي، ويمكن برهان ذلك فيما يلي:

يأخذ نموذج الانحدار الخطي المتعدد الصيغة التالية:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

وإن:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\mathbf{y}$$

وبالتعويض نحصل على:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\left(\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right) = \left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\mathbf{x}\boldsymbol{\beta} + \left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\boldsymbol{\varepsilon}$$

وبأخذ القيمة المتوقعة لطرفي المعادلة نحصل على:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

تباين ($\hat{\boldsymbol{\beta}}$):

يمكن إيجاد تباين ($\hat{\boldsymbol{\beta}}$) كما يلي:

$$\operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right) = E\left\{\left[\hat{\boldsymbol{\beta}} - E\left(\hat{\boldsymbol{\beta}}\right)\right]\left[\hat{\boldsymbol{\beta}} - E\left(\hat{\boldsymbol{\beta}}\right)\right]^{T}\right\}$$

$$\begin{aligned}\operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right) &= E\left\{\left[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right]\left[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right]^{T}\right\}\\ &= E\{(\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\varepsilon\varepsilon^{T}\mathbf{x}(\mathbf{x}^{T}\mathbf{x})^{-1}\} \qquad (7\text{-}62)\\ &= (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}E(\varepsilon\varepsilon^{T})\mathbf{x}(\mathbf{x}^{T}\mathbf{x})^{-1}\\ &= (\mathbf{x}^{T}\mathbf{x})^{-1}\mathbf{x}^{T}\sigma^{2}\Omega\mathbf{x}(\mathbf{x}^{T}\mathbf{x})^{-1}\end{aligned}$$

حيث إن Ω هي المصفوفة المعرفة حسب المعادلة رقم (7.60).

وتوضح هذه المعادلة أن تباين معلمات نموذج الانحدار المقدرة في حالة وجود ارتباط ذاتي من الرتبة الأولى تختلف من تباينها من حالة عدم وجود ارتباط ذاتي، أي أن:

$$\operatorname{var}\left(\hat{\boldsymbol{\beta}}\right)_{AR(1)} = \left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}\sigma^{2}\Omega\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1} \neq \operatorname{var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^{2}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}$$

ويترتب على هذا الاختلاف أن تكون النتائج التي نحصل عليها من الإحصاء الاستدلالي (إحصاءات t و F وفترات الثقة والتنبؤ) غير صحيحة.

ويمكن تبيان الاختلاف بين تباين مقدرات المربعات الصغرى في ظل وجود ارتباط ذاتي بين حدود الخطأ وعدمه، بحساب تباين معامل الانحدار الخطي البسيط في صيغة الانحرافات، حيث يأخذ النموذج الصيغة التالية:

$$y_{t} = \beta x_{t} + \varepsilon_{t}$$

وأن الارتباط بين حدود الخطأ من الرتبة الأولى يأخذ الصيغة التالية:

$$\varepsilon_t = \rho\varepsilon_{t-1} + a_t$$

والآن لإيجاد تباين معامل الانحدار ($\hat{\boldsymbol{\beta}}$) نستخدم الصيغة التالية:

$$\operatorname{var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2 \left(\mathbf{x}^T\mathbf{x}\right)^{-1} \mathbf{x}^T\Omega\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}$$

وبما أن:

$$(\mathbf{x}^T\mathbf{x}) = \sum_{t=1}^{N} x_t^2 \text{ و } \left(\mathbf{x}^T\mathbf{x}\right)^{-1} = \frac{1}{\sum_{t=1}^{N} x_t^2}$$

فإن:

$$\operatorname{var}(\hat{\beta})_{AR(1)} = \frac{\sigma^2}{\sum_{t=1}^{N} x_t^2}(x_1\ x_2\ ...\ x_N)\begin{pmatrix} 1 & \rho & \rho^2 & ... & \rho^{N-1} \\ \rho & 1 & \rho & ... & \rho^{N-2} \\ \rho^2 & \rho & 1 & ... & \rho^{N-3} \\ . & . & . & ... & . \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & ... & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_N \end{pmatrix} \cdot \frac{1}{\sum_{t=1}^{N} x_t^2}$$

$$\operatorname{var}(\hat{\beta})_{AR(1)} = \frac{\sigma^2}{\sum_{t=1}^{N} x_t^2}\left(1 + \frac{2\rho\sum_{t=1}^{N-1} x_t x_{t+1}}{\sum_{t=1}^{N} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{N-2} x_t x_{t+2}}{\sum_{t=1}^{N} x_t^2} + ... + \frac{2\rho^{N-1} x_1 x_N}{\sum_{t=1}^{N} x_t^2}\right)$$

ويلاحظ أن تباين ($\hat{\beta}$) في ظل الارتباط الذاتي يختلف من تباينه في حالة عدم وجود ارتباط ذاتي بين حدود الخطأ الذي يأخذ الصيغة التالية:

$$\operatorname{var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2}$$

كما يلاحظ أن تباين ($\hat{\beta}$) في حالة وجود ارتباط ذاتي يساوي تباين ($\hat{\beta}$) في حالة عدم وجود ارتباط ذاتي زائداً حد يعتمد على معامل الارتباط الذاتي ρ وعلى التغاير بين قيم المتغير المستقل (X). فإذا كانت قيمة الارتباط الذاتي مساوية للصفر نجد أن:

$$\operatorname{var}\left(\hat{\beta}\right) = \operatorname{var}\left(\hat{\beta}\right)_{AR(1)}$$

أما إذا كانت قيمة الارتباط الذاتي أكبر من الصفر فإن:

$$\operatorname{var}\left(\hat{\beta}\right)_{AR(1)} > \operatorname{var}\left(\hat{\beta}\right)$$

٧-٤-٥ بعض الطرق المستخدمة في الكشف عن الارتباط الذاتي:

تطرقنا في الفصل الثالث إلى بعض الطرق البيانية للكشف عن الارتباط الذاتي باستخدام رسم انتشار البواقي المعيارية مع الزمن أو البواقي المعيارية المبطئة لفترة زمنية واحدة. وفي هذا الجزء سنتعرف على بعض الطرق التحليلية الشائعة الاستخدام في الكشف عن هذه المشكلة.

٧-٤-٥-١ الرسم البياني لمعاملات الارتباط الذاتي مع الفجوات الزمنية:

من الطرق التي تستخدم للكشف عن وجود الارتباط الذاتي هو الرسم البياني لمعاملات الارتباط الذاتي بين البواقي مع الفجوات الزمنية (lags) الذي يعرف باسم (Correlogram). حيث يتم أولاً حساب معاملات العينة للارتباط الذاتي للبواقي الاعتيادية ($\hat{\rho}_s$) كما يلي:

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^{n} e_t e_{t-s}}{\sum_{t=1}^{n} e_t^2}, \qquad s = 1,2,\ldots,m \qquad (7\text{-}63)$$

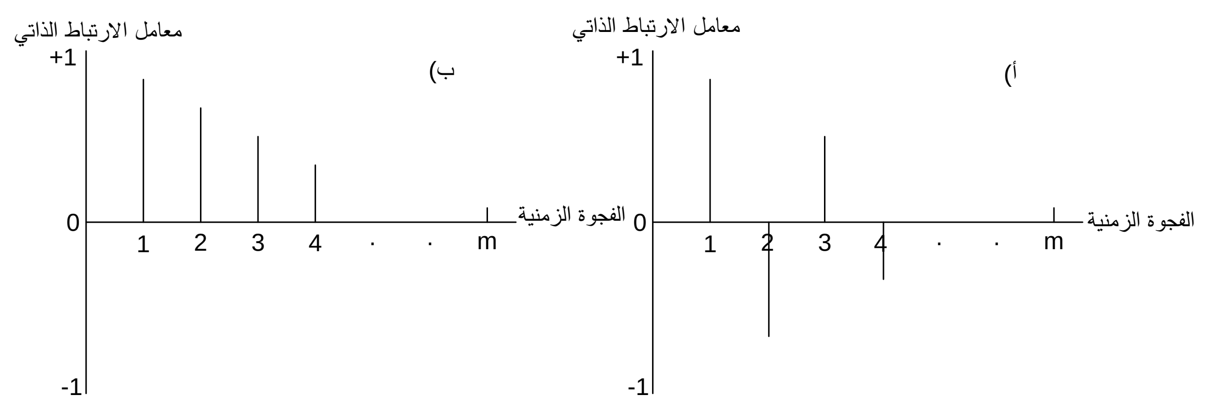
حيث إن:

e_t = الباقي عند الفترة t.

s= الفجوة الزمنية التي تأخذ القيم من ١ إلى m، ويقترح أن لا تزيد قيمة m عن ربع عدد المشاهدات، أي:

$$m \le \frac{n}{4}$$

ومن ثم يتم رسم شكل معاملات الارتباط المقدرة مع طول الفجوة. فإذا بدا لنا من شكل الــ Correlogram أن قيم معاملات الارتباط الذاتي تتناقص ببطء وتقترب من الصفر بزيادة طول الفجوة دلّ ذلك على وجود ارتباط ذاتي من الرتبة الأولى AR(1) (شكل رقم (٧-٩). وتعد قيم معاملات الارتباط الذاتي للبواقي التي تقع خارج فترة الثقة ($\pm 1.96/\sqrt{n}$)-وهي فترة ثقة ٩٥%- تختلف معنوياً عن الصفر عند مستوى معنوية ٥%.



شكل رقم (٧-٩): رسم معاملات الارتباط الذاتي مع الفجوة الزمنية - حالتان من وجود ارتباط ذاتي بين بواقي نموذج انحدار خطي

٧-٤-٥-٢ اختبار ديربن-واتسون (Durbin -Watson Test):

من الاختبارات المستخدمة بشكل واسع للكشف عن الارتباط الذاتي من الرتبة الأولي هو اختبار ديربن–واتسون والمتضمن في معظم برامج الإحصاء الجاهزة. ويستخدم إحصاء ديربن-واتسون لاختبار ثلاثة فروض هي:

- وجود ارتباط ذاتي موجب:

فرض العدم: $H_0 : \rho = 0$ في مقابل الفرض البديل: $H_1 : \rho > 0$

-وجود ارتباط ذاتي سالب:

فرض العدم: $H_0 : \rho = 0$ في مقابل الفرض البديل: $H_1 : \rho < 0$

-وجود ارتباط ذاتي سالب أو موجب (اختبار ذو طرفين):

فرض العدم: $H_0 : \rho = 0$ في مقابل الفرض البديل: $H_1 : \rho \neq 0$

إحصاء ديربن-واتسون:

يأخذ إحصاء ديربن-واتسون (Durbin & Watson (1951), pp.159-178) الصيغة التالية:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \qquad (7\text{-}64)$$

حيث إن:

e_t = باقي نموذج الانحدار الخطي الموفق للفترة t، أي أن: $e_t = y_t - \hat{y}_t$

e_{t-1} = الباقي للفترة السابقة لـ t ‏(t-1).

n = عدد المشاهدات.

ويلاحظ أن إحصاء d عبارة نسبة مجموع مربعات الفروق بين البواقي المتتالية إلى مجموع مربعات البواقي (RSS). ولاستخدام إحصاء ديربن-واتسون لا بد من توفر الشروط التالية:

- أن يكون الارتباط بين حدود الخطأ من الرتبة الأولى AR(1).
- أن يحتوي نموذج الانحدار على المعامل الثابت (Intercept).
- ألا يتضمن نموذج الانحدار متغيرات متباطئة (lagged variables) كمتغيرات مفسرة. فمثلاً لا يصلح الاختبار في حالة النموذج التالي:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 y_{t-1} + \varepsilon_t$$

لأن النموذج يضم المتغير المتباطئ Y_{t-1} ضمن المتغيرات المستقلة.

- ألا توجد مشاهدات مفقودة (missing observations) سواء في المتغير التابع أو في المتغيرات المستقلة. فمثلاً لا يصلح استخدام إحصاء ديربن-واتسون إذا كانت لدينا مشاهدات مفقودة لسنة أو بضعة سنوات لبيانات سلسلة زمنية تمتد من ١٩٧٠ إلى ١٩٩٥م.

العلاقة بين إحصاء d ومعامل الارتباط الذاتي المقدر $(\hat{\rho})$:

بفك قوس بسط المعادلة (7.64) نحصل على:

$$d = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

ونظراً لتقارب القيم التالية بحيث إن:

$$\sum_{t=2}^{n} e_t^2 \approx \sum_{t=2}^{n} e_{t-1}^2 \approx \sum_{t=1}^{n} e_t^2$$

فإنه يمكن كتابة إحصاء d كما يلي:

$$d \approx \frac{2\sum_{t=1}^{n} e_t^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \approx 2\left(1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right) \approx 2(1-\hat{\rho}) \qquad (7\text{-}65)$$

ومن المعادلة (7.65) أمكن الوصول إلى النتائج التالية:

- تنحصر قيم d ما بين قيمتي الصفر و(٤)، أي:

$$0 \le d \le 4$$

ذلك لأن قيمة ($\hat{\rho}$) تتراوح ما بين سالب واحد وموجب واحد.

- إذا كانت قيمة معامل الارتباط مساوية للصفر ($\hat{\rho} = 0$) فإن قيمة d=0 ودلّ ذلك على عدم وجود ارتباط ذاتي.
- إذا كانت قيمة معامل الارتباط تساوي سالب واحد ($\hat{\rho} = -1$) فإن قيمة d=4، ودلّ ذلك على وجود ارتباط ذاتي سالب تام بين القيم المتتالية للبواقي.
- إذا كانت قيمة معامل الارتباط تساوي موجب واحد صحيح، دلّ ذلك على وجود ارتباط ذاتي موجب تام بين البواقي المتتالية وتصبح قيمة d=0.

قاعدة القرار:

لقد أعد كل من ديربن-واتسون جداول لاستخدامها للكشف عن وجود الارتباط الذاتي* حيث تحدد قيم d الجدولية وفقاً لثلاثة عوامل هي:

- عدد المشاهدات (n).
- عدد المتغيرات المستقلة (p).
- مستوي المعنوية (α).

وتوجد قيمتين لـ d بالجدول هما:

- الحد الأعلى (Upper bound) ويرمز له بـ d_U.
- الحد الأدنى (Lower bound) ويرمز له بـ d_L.

ويتم اتخاذ القرار حول وجود ارتباط ذاتي حسب قاعدة القرار الموضحة بالجدول رقم (٧-١٠) والشكل رقم (٧-١٠).

* يتم في بعض البرامج الإحصائية حساب قيمة الاحتمال (المعنوية) لاختبار ديربن - واتسون، مثل برنامج **SHAZAM**

جدول رقم (٧-١٠): اختبار ديربن-واتسون - قاعدة القرار

فرض العدم	القرار	الشرط
لا يوجد ارتباط ذاتي موجب	رفض	$0 < d < d_L$
لا يوجد ارتباط ذاتي موجب	عدم قرار	$d_L \le d \le d_U$
لا يوجد ارتباط ذاتي سالب	رفض	$4-d_L < d < 4$
لا يوجد ارتباط ذاتي سالب	عدم قرار	$4-d_U \le d \le 4-d_L$
لا يوجد ارتباط ذاتي (سالب أو موجب)	قبول	$d_U < d < 4-d_U$

رفض فرض العدم (يعني وجود ارتباط ذاتي موجب)	منطقة عدم قرار	لا يوجد دليل لرفض فرض العدم (يعني عدم وجود ارتباط ذاتي)	منطقة عدم قرار	رفض فرض العدم (يعني وجود ارتباط ذاتي سالب)

0 d_L d_U $4-d_U$ $4-d_L$ 4

شكل رقم (٧-١٠): اختبار ديربن-واتسون - مناطق القبول والرفض

ومن عيوب اختبار ديربن-واتسون هو وجود منطقة عدم اتخاذ القرار (Indecisive zone) إذ لا يمكننا الحكم بوجود أو عدم وجود ارتباط ذاتي.

٧-٤-٥-٣ اختبار بروش-جودفري (Breusch-Godfrey):

يعاب على اختبار ديربن-واتسون أنه يستخدم للكشف عن الارتباط الذاتي من الرتبة الأولى فقط. ففي حال وجود ارتباط من الرتبة الأولى أو أعلى تستخدم اختبارات أخرى من بينها اختبار بروش-جودفري. واختبار بروش-جودفري اختبار عام يستخدم للكشف عن الارتباط الذاتي من أي رتبة كرتبة AR(1) وAR(2) ... إلخ بالإضافة إلى كشف الارتباط الذاتي حتى في حالة وجود متغيرات متباطئة للمتغير التابع $(Y_{t-1}, Y_{t-2},..)$ كمتغيرات مفسرة. وفرض العدم لهذا الاختبار: $H_0 : \rho_1 = \rho_2 = ... = \rho_q = 0$ مقابل فرض البديل $H_0 : \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 ... \text{ or } \rho_q \neq 0$.

وفيما يلي خطوات اختبار بروش - جودفري (Brooks, 2002;pp. 164-165):

١- بناء نموذج الانحدار الخطي باستخدام طريقة المربعات الصغرى $y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_{pt} x_{pt} + \varepsilon_t$ بافتراض أن حد الخطأ ε_t يتبع الانحدار الذاتي من الرتبة q AR(q) كما يلي:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + ... \rho_q \varepsilon_{t-q} + a_t$$

حيث إن a_t حد خطأ عشوائي (White noise error) يتبع التوزيع الطبيعي بوسط حسابي مساو للصفر وتباين ثابت مساو σ^2 $a_t \sim N\left(0,\sigma^2\right)$

٢- الحصول على معامل التحديد R^2 من نموذج انحدار البواقي على المتغيرات المستقلة والبواقي المبطأة لعدد q فترة كما يلي:

$$e_t = \beta_0+\beta_1 x_{1t}+\ldots+\beta_{pt} x_{pt}+\gamma_1 e_{t-1}+\gamma_2 e_{t-2}+\ldots+\gamma_q e_{t-q}+a_t \qquad (7\text{-}66)$$

٣- إذا كان حجم العينة (n) كبيراً، فإن الإحصاءة• $(n-q)R^2$ تتبع توزيع مربع كاي بدرجات مساوية لرتبة الانحدار الذاتي (q)، أي $(n-q)R^2 \sim \chi^2_q$. فإذا كانت قيمة الإحصاءة $(n-q)R^2$ أكبر من القيمة الحرجة لتوزيع مربع كاي بدرجات حرية q ومستوى معنوية محدد ($\chi^2_{\alpha,q}$) فيتم رفض فرض العدم، أي أن هناك ارتباطاً ذاتياً بين بواقي النموذج رتبته q .

٧-٤-٥-٤ اختبار التلاحقات (Runs Test):

من الطرق المستخدمة للكشف عن الارتباط الذاتي اختبار التلاحقات . ويهدف هذا الاختبار إلى الكشف عما إذا كان ترتيب البواقي عشوائياً أو يتخذ نمطاً معيناً لا يعزى للصدفة وحدها. ولتوضيح الاختبار ننظر إلى إشارات البواقي (+، -)، فمثلاً إذا كان لدينا قيم البواقي التالية لنموذج ما:

الباقي	١,٥-	٢,٢-	١,٥+	١,٧+	٢,٦+	٣,٠-	٢,٠-	٤,٠+	٢,٦+
الإشارة	-	-	+	+	+	-	-	+	+

فيقال إن البواقي تشتمل على (٤) تلاحقات، تتألف التلاحقة الأولى من إشارتين سالبتين ونقول إن طولها اثنان، وتتألف التلاحقة الثانية من ثلاثة إشارات موجبة ونقول إن طولها ثلاثة وتتألف الثالثة من إشارتين سالبتين وطولها اثنان والتلاحقة الرابعة والأخيرة تتألف من إشارتين موجبتين ويبلغ طولها اثنين.

افترض أن:

N = عدد المشاهدات.

N_1 = عدد الإشارات الموجبة (عدد مشاهدات البواقي الموجبة).

N_2 = عدد الإشارات السالبة (عدد مشاهدات البواقي السالبة).

n= عدد التلاحقات.

• يتم في بعض البرامج كنظام ساس SAS وEviews استبدال القيم المفقودة بسبب ادخال البواقي كمتغيرات متباطئة بقيمة الصفر ومن ثم يتم حساب الإحصاءة nR^2 (انظر Johnston and Dinardo, 1997, pp. 183-184) بدلاً من $(n-q)R^2$

ففي المثال السابق نجد أن:

$N = ٩$ ،$N_1 = ٥$ ، $N_2 = ٤$ ، $n = ٤$.

والآن بافتراض أن النتائج المتلاحقة (البواقي) مستقلة وأن كل من N_1 و N_2 أكبر من ١٠، فإن عدد التلاحقات يتوزع تقريباً حسب التوزيع الطبيعي وسطه الحسابي وتباينه كالآتي:

الوسط الحسابي:

$$E(n)=\frac{2N_1N_2}{N_1+N_2}+1 \qquad (7\text{-}67)$$

والتباين:

$$\sigma_n^2=\frac{2N_1N_2(2N_1N_2-N_1-N_2)}{(N_1+N_2)^2(N_1+N_2-1)} \qquad (7\text{-}68)$$

إذا كان الفرض الصفري هو أن العينة عشوائية والفرض البديل هو العكس، فإننا نستخدم اختباراً ذا ذيلين ويتم رفض الفرض الصفري عند مستوى معنوية ٥% إذا وقع عدد التلاحقات (n) خارج المنطقة:

$$E(n)\pm 1.96\sigma_n \qquad (7\text{-}69)$$

إذا كان أحد العددين N_1 و N_2 أو كلاهما أقل من ١٠ فلا يجوز التقريب بالتوزيع الطبيعي؛ وفي هذه الحالة تستخدم جداول خاصة لإجراء الاختبار (انظر مثلاً Gujarati and Porter, 2009)

٧-٤-٦ بعض طرق معالجة الارتباط الذاتي من الدرجة الأولى:

٧-٤-٦-١ طريقة المربعات الصغرى المعممة:

في حالة وجود ارتباط ذاتي من الرتبة الأولى AR(1) فإن مصفوفة تباين-تغاير حد الخطأ تأخذ الصيغة التالية:

$$E(\varepsilon\varepsilon^T)=\sigma^2\Omega=\sigma^2\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ . & . & . & \dots & . \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix}$$

ويلاحظ أن مصفوفة تباين - تغاير الخطأ تعتمد فقط على معلمتين هما ρ و σ^2. فإذا تم معرفة قيم هاتين المعلمتين فإنه يمكن استخدام طريقة الصغرى المعممة (Generalized Least Squares (GLS)) وذلك بتحويل النموذج الأصلي إلى نموذج مصحح؛ حيث يتم تحويل قيم كل من المتغير التابع والمتغيرات المفسرة ومن ثم يجري حل نموذج الانحدار بطريقة المربعات الصغرى الاعتيادية. وتتلخص طريقة المربعات الصغرى المعممة (GLS) في الآتي:

يأخذ نموذج الانحدار الخطي المتعدد الصيغة التالية:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

وفي ظل الارتباط الذاتي نجد أن: $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ و $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\boldsymbol{\Omega}$

ولاستخدام طريقة المربعات الصغرى المعممة يتم تصحيح النموذج الأصلي وذلك بضرب قبلي لطرفي معادلة نموذج الانحدار الخطي المتعدد بالمصفوفة $\Omega^{-1/2}$ كما يلي:

$$\Omega^{-1/2}\mathbf{y} = \Omega^{-1/2}\mathbf{x}\boldsymbol{\beta} + \Omega^{-1/2}\boldsymbol{\varepsilon}$$

حيث يتم حساب المصفوفة $\Omega^{-1/2}$ على النحو التالي:

وبما أن:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ . & . & . & \dots & . \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix}$$

فإن معكوس هذه المصفوفة هو:

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ . & . & . & \dots & . & . \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

وإن المصفوفة $\Omega^{-1/2}$ تأخذ الصيغة التالية:

$$\Omega^{-1/2}=\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}$$

ويتم تحويل المتغيرات على النحو التالي:

$$\mathbf{y}^*=\Omega^{-1/2}\mathbf{y}$$

$$\mathbf{y}^*=\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 \\ . & . & . & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ y_N \end{pmatrix}=\begin{pmatrix} \sqrt{1-\rho^2}\,y_1 \\ y_2-\rho y_1 \\ . \\ . \\ . \\ y_N-\rho y_{N-1} \end{pmatrix}$$

و

$$\mathbf{x}^*=\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 \\ . & . & . & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}\begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{p1} \\ 1 & x_{12} & x_{22} & \ldots & x_{p2} \\ 1 & x_{13} & x_{23} & \ldots & x_{p3} \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ 1 & x_{1N} & x_{2N} & \ldots & x_{pN} \end{pmatrix}=\begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\,x_{11} & \ldots & \sqrt{1-\rho^2}\,x_{p1} \\ 1-\rho & x_{12}-\rho x_{11} & \ldots & x_{p2}-\rho x_{p1} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ 1-\rho & x_{1N}-\rho x_{1,N-1} & \ldots & x_{pN}-\rho x_{p,N-1} \end{pmatrix}$$

و

$$\boldsymbol{\varepsilon}^* = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ . & . & . & \dots & 0 & 0 \\ . & . & . & \dots & . & . \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \cdot \\ \cdot \\ \varepsilon_N \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}\,\varepsilon_1 \\ \varepsilon_2 - \rho\varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_N - \rho\varepsilon_{N-1} \end{pmatrix}$$

ويلاحظ أن القيمة المتوقعة لحد الخطأ للنموذج المصحح مساوية للصفر، أي أن:

$$E(\varepsilon^*) = E(\Omega^{-1/2}\varepsilon) = \mathbf{0}$$

وأن مصفوفة تباين - تغاير حدود الخطأ تأخذ الصيغة التالية:

$$E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*T}) = E(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{T}\boldsymbol{\Omega}^{-1/2T}) = \boldsymbol{\Omega}^{-1/2}\sigma^2\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1/2T} = \sigma^2\mathbf{I}_{NxN}$$

والآن يمكن استخدام طريقة المربعات الصغرى الاعتيادية للمتغيرات المحولة كالآتي:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{x}^{*T}\mathbf{x}^*)^{-1}\mathbf{x}^*\mathbf{y}^* = (\mathbf{x}^{T}\boldsymbol{\Omega}^{-1}\mathbf{x})^{-1}\mathbf{x}^{T}\boldsymbol{\Omega}^{-1}\mathbf{y} \qquad (7\text{-}70)$$

كما يتم إيجاد مصفوفة تباين - تغاير حد الخطأ كما يلي:

$$E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*T}) = \sigma^2(\mathbf{X}^{*T}\mathbf{X}^*)^{-1} = \sigma^2(\mathbf{X}^{T}\Omega^{-1}\mathbf{X})^{-1} \qquad (7\text{-}71)$$

٧-٤-٦-٢ طريقة الفروق المعممة (Generalized Difference Method):

يمكن الحصول على نفس نتائج طريقة المربعات الصغرى باستخدام طريقة الفروق المعممة. وتعتمد هذه الطريقة أيضاً على معرفة قيمة معامل الارتباط الذاتي (ρ). وتتلخص هذه الطريقة في الخطوات التالية:

- يأخذ نموذج الانحدار الخطي المتعدد الصيغة التالية:

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_p x_{pt} + \varepsilon_t \qquad (7\text{-}72)$$

وبما أن:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad u_t \sim NID(0,\ \sigma_u^2)$$

فإنه يمكن كتابة المعادلة (7.72) كما يلي:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_p x_{pt} + \rho\varepsilon_{t-1} + u_t \qquad (7\text{-}73)$$

وللمشاهدة السابقة (t-1) يمكن كتابة المعادلة (7.73) على النحو التالي:

$$y_{t-1}=\beta_0+\beta_1 x_{1,t-1}+\ldots+\beta_p x_{p,t-1}+\varepsilon_{t-1} \qquad (7\text{-}74)$$

وبضرب طرفي المعادلة (7.74) بمعامل الارتباط الذاتي (ρ) نحصل على:

$$\rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{1,t-1} + \ldots + \rho\beta_p x_{p,t-1} + \rho\varepsilon_{t-1} \qquad (7\text{-}75)$$

وبطرح المعادلة (7.75) من المعادلة (7.74) نحصل على:

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_{1t} - \rho x_{1,t-1}) + \ldots + \beta_p(x_{pt} - \rho x_{p,t-1}) + u_t \quad t = 2,3,\ldots,n \quad (7.76a)$$

أو

$$y^*_t = \beta^*_0 + \beta_1 x^*_{1t} + \ldots + \beta_p x^*_{pt} + u_t \qquad t = 2,3,\ldots,n \qquad (7.76b)$$

حيث إن

$$y^*_t = y_t - \rho y_{t-1} ,$$
$$\beta^*_0 = \beta_0(1-\rho)$$
$$x^*_{1t} = (x_{1t} - \rho x_{1,t-1}), \ldots, x^*_{p,t} = (x_{p,t} - \rho x_{p,t-1}),$$
$$u_t \sim N(0,\sigma^2_u)$$

ويتمثل الفرق الوحيد بين طريقة الفروق المعممة وطريقة المربعات الصغرى المعممة (GLS) هو أننا نفقد المشاهدة الأولى باستخدام الطريقة الأولى. ولتفادي ضياع المشاهدة الأولى في عملية الفروق، يتم تقدير المشاهدة لكل من المتغير التابع والمتغيرات المفسرة كما يلي:

$$y^*_1 = y_1\sqrt{1-\rho^2}$$
$$x^*_{h,1} = x_{h,1}\sqrt{1-\rho^2} \qquad h = 1,2,\ldots,p$$

ويلاحظ أن هاتين الطريقتين تعتمدان على معرفة قيمة معامل الارتباط الذاتي (ρ). ولكن في الواقع العملي نادراً ما تكون قيمة معامل الارتباط الذاتي معلومة. ومن ثم نحتاج لتقديرها حتى يتسنى لنا استخدام أي من الطريقتين. ولحسن الحظ يوجد عدد من الطرق المستخدمة لتقدير قيمة معامل الارتباط الذاتي (ρ). فيما يلي نستعرض بعض هذه الطرق:

أ- تقدير قيمة معامل الارتباط الذاتي باستخدام إحصاء ديربن-واتسون:

كما أوضحنا من قبل أن العلاقة بين معامل الارتباط الذاتي وإحصاء ديربن-واتسون تأخذ الصيغة التالية:

$$d \approx 2\left(1-\hat{\rho}\right)$$

وبإعادة تنظيم هذه المعادلة نجد أن:

$$\hat{\rho} \approx 1-0.5\mathrm{d} \qquad (7\text{-}77)$$

ومن ثم يتم استبدال قيمة $\left(\rho\right)$ الواردة في مصفوفة التصحيح أو الواردة في معادلة الفروق المعممة بقيمة معامل الارتباط المقدرة حسب المعادلة (7.71) وإجراء حل النموذج بطريقة المربعات الصغرى الاعتيادية.

ب- تقدير معامل الارتباط الذاتي باستخدام طريقة ثيل - نجار (Theil-Nagar):

يستخدم تقدير معامل الارتباط الذاتي من إحصاء ديربن-واتسون في حالة العينات الكبيرة (n>30). وأما في حالة العينات الصغيرة، اقترح كل من ثيل ونجار (Theil & Nagar, (1961) pp.793-806) العلاقة التالية لتقدير قيمة $\left(\rho\right)$.

$$\hat{\rho}=\frac{n^2(1-0.5d)+p^2}{n^2-p^2} \qquad (7\text{-}78)$$

حيث إن:

n = عدد المشاهدات.

d = إحصاء ديربن-واتسون.

p =عدد المتغيرات المستقلة في نموذج الانحدار.

ويتم استبدال قيمة معامل الارتباط الذاتي في مصفوفة التصحيح $\left(\Omega^{-1/2}\right)$ أو معادلة الفروق المعممة بالقيمة المقدرة لها $\left(\hat{\rho}\right)$ ومن ثم يتم حل معادلة نموذج الانحدار.

ج- طريقة كوكرين - أوركت (Cochrane-Orcutt Method) لتقدير قيمة معامل الارتباط الذاتي:

تعد هذه الطريقة من أكثر الطرق استخداماً لتقدير قيمة معامل الارتباط الذاتي بين حدود الخطأ ومتضمنة في معظم برامج الإحصاء الجاهزة. حيث اقترح كل من كوكرين وأوركت (Cochrane and Orcutt (1949), pp.32-61) الخطوات التالية لتقدير قيمة $\left(\rho\right)$.

١. يتم أولاً تقدير معالم النموذج الخطي بطريقة المربعات الصغرى الاعتيادية حسب المعادلة التالية:

$$y_t=\beta_0+\beta_1 x_{1t}+....+\beta_p x_{pt}+\varepsilon_t$$

٢. يتم حساب بواقي النموذج الموفق ($e_t = y_t - \hat{y}_t$).

٣. يتم بناء نموذج البواقي التالي:

$$e_t = \hat{\rho} e_{t-1} + v_t$$

للحصول على تقدير معامل الارتباط الذاتي بين حدود الخطأ المتتالية.

٤. تستخدم قيمة معامل الارتباط الذاتي المقدرة ($\hat{\rho}$) لتوفيق معادلة الفروق المعممة التالية:

$$(y - \hat{\rho} y_{t-1}) = \beta_0 (1 - \hat{\rho}) + \beta_1 (x_{1t} - \hat{\rho} x_{1,t-1}) + ... + \beta_p (x_{p,t} - \hat{\rho} x_{p,t-1}) + u_t$$

أو

$$y^*_t = \beta^*_0 + \beta_1 x^*_{1t} + ... + \beta_p x^*_{pt} + u_t$$

حيث إن:

$$x^*_{ht} = (x_{ht} - \hat{\rho} x_{h,t-1}) \text{ for } h = 1, 2, ..., p \text{ ؛ } \beta^*_0 = \beta_0 (1 - \hat{\rho}) \text{ ؛ } y^*_t = (y_t - \hat{\rho} y_{t-1})$$

٥. بما أنه لا توجد معلومة توضح أن قيمة معامل الارتباط الذاتي المقدرة في الخطوة الثالثة هي أفضل مقدر للقيمة الحقيقية للمعامل (ρ)، يتم استخدام قيم معاملات النموذج المقدرة في الخطوة السابقة في نموذج الانحدار الأصلي ومن ثم يتم الحصول على بواقي جديدة. وباستخدام البواقي الجديدة يتم إجراء النموذج التالي:

$$e^{**}_t = \hat{\rho} e^{**}_{t-1} + w_t$$

للحصول على تقدير جديد لقيمة معامل الارتباط الذاتي ($\hat{\hat{\rho}}$) ويستخدم المقدر الجديد لتوفيق معادلة الفروق المعممة حسب الخطوة الرابعة.

٦. وبما أننا أيضاً لا نعلم أن قيمة معامل الارتباط الذاتي المقدرة في الجولة الثانية أفضل مقدر لقيمة المعلمة الحقيقية، نستمر في بناء نموذج البواقي مرة ثالثة. وهكذا تستمر العملية إلى الحد الذي تصبح فيه الفروق بين قيم معامل الارتباط الذاتي المقدرة صغيرة جداً. ومن ثم يتم استخدام التقدير النهائي للمعامل في معادلة الفروق المعممة لتقدير معالم النموذج والإحصاءات الأخرى.

د- طريقة (Hildreth-Lu) لتقدير قيمة معامل الارتباط الذاتي بين حدود الخطأ المتتالية:

تعتمد هذه الطريقة على بناء عدة نماذج باستخدام طريقة المربعات الصغرى المعممة (GLS) أو معادلة الفروق المعممة، وذلك باختيار قيم تقديرية مختلفة لمعامل الارتباط الذاتي. ويقترح كل من Hildreth وLu اختيار قيم معامل الارتباط الذاتي التي تحقق أقل قيمة لمجموع مربعات البواقي (RSS). ويعاب على هذه الطريقة أنها تحتاج إلى

حسابات مطولة، ولحسن الحظ توجد بعض برامج الإحصاء الجاهزة تتضمن هذه الطريقة لحساب قيمة معامل الارتباط الذاتي $\hat{\rho}$.

هـ- طريقة ديربن ذات المرحلتين لتقدير معامل الارتباط الذاتي بين حدود الخطأ المتتالية:

اقترح ديربن (Durbin (1960) pp.139-153) طريقة أخرى لتقدير معامل الارتباط الذاتي بين حدود الخطأ المتتالية تعرف بطريقة ديربن ذات المرحلتين (Durbin's two-step method for estimating ρ). حيث تمت الاستفادة في هذه الطريقة من معادلة الفروق المعممة:

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_{1t} - \rho x_{1,t-1}) + \ldots + \beta_p(x_{pt} - \rho x_{p,t-1}) + u_t \quad t = 2,3,\ldots n$$

وبإعادة تنظيم هذه المعادلة نحصل على:

$$\begin{aligned} y_t &= \beta_0(1-\rho) + \beta_1(x_{1t} - \rho x_{1,t-1}) + \ldots + \beta_p(x_{pt} - \rho x_{p,t-1}) + \rho y_{t-1} + u_t \\ y_t &= \beta_0(1-\rho) + \beta_1 x_{1t} - \beta_1 \rho x_{1,t-1} + \ldots + \beta_p x_{pt} - \beta_p \rho x_{p,t-1} + \rho y_{t-1} + u_t \quad t = 2,3,\ldots,n \end{aligned} \tag{7-79}$$

ويلاحظ من هذه المعادلة أن معامل المتغير المتباطئ (Y_{t-1}) هو معامل الارتباط الذاتي ($\hat{\rho}$). وبحل هذه المعادلة نحصل على تقدير معامل الارتباط ($\hat{\rho}$). وتتلخص هذه الطريقة في التالي:

- في الخطوة الأولى يتم إجراء انحدار المتغير التابع على قيمته المبطئة لفترة زمنية واحدة (Y_{t-1}) وعلى المتغيرات المستقلة ($X_{1t}, \quad X_{2t}, \quad \ldots, X_{pt}$) وعلى قيم المتغيرات المستقلة المبطئة لفترة زمنية واحدة ($X_{1,t-1}$, $X_{2,t-1}$, $\ldots$,$X_{p,t-1}$). ويعد قيمة معامل المتغير المتباطئ (Y_{t-1}) التي نحصل عليها هي مقدر معامل الارتباط الذاتي بين حدود الخطأ المتتالية.
- في الخطوة الثانية يتم استخدام مقدر معامل الارتباط الذاتي في حل نموذج الفروق المعممة التالية:

$$y_t - \hat{\rho} y_{t-1} = \beta_0(1-\hat{\rho}) + \beta_1(x_{1t} - \hat{\rho} x_{1,t-1}) + \ldots + \beta_p(x_{pt} - \hat{\rho} x_{p,t-1}) + u_t$$

٧-٤-٧ مثال:

يوضح الجدول رقم (٧-١١) بيانات الواردات والناتج المحلي الإجمالي الخاصة بالمملكة العربية السعودية للفترة من عام ١٩٦٩ إلى عام ١٩٩٦م. المطلوب تقدير دالة الطلب على الواردات، وذلك بإجراء انحدار الواردات على الناتج المحلي الإجمالي واختبار وجود مشكلة الارتباط الذاتي من الدرجة الأولى ومعالجتها إن وجدت.

جدول رقم (٧-١١): الواردات والناتج المحلي الإجمالي للمملكة العربية السعودية

العام	الواردات (بليون ريال)	الناتج المحلي الإجمالي (بليون ريال)
1969	3.38	17.15
1970	3.20	22.58
1971	3.67	27.86
1972	4.71	40.09
1973	7.31	98.84
1974	10.15	139.23
1975	14.82	163.89
1976	30.69	203.94
1977	51.66	223.82
1978	69.18	247.62
1979	82.22	383.59
1980	100.35	517.99
1981	119.30	522.18
1982	139.34	411.80
1983	135.42	368.40
1984	118.74	347.42
1985	85.56	310.03
1986	70.78	267.85
1987	75.31	272.00
1988	81.58	276.91
1989	79.22	304.08
1990	90.28	384.99
1991	108.93	435.04
1992	124.61	452.30
1993	105.62	434.57
1994	87.45	441.74
1995	105.19	461.62
1996	103.98	500.93

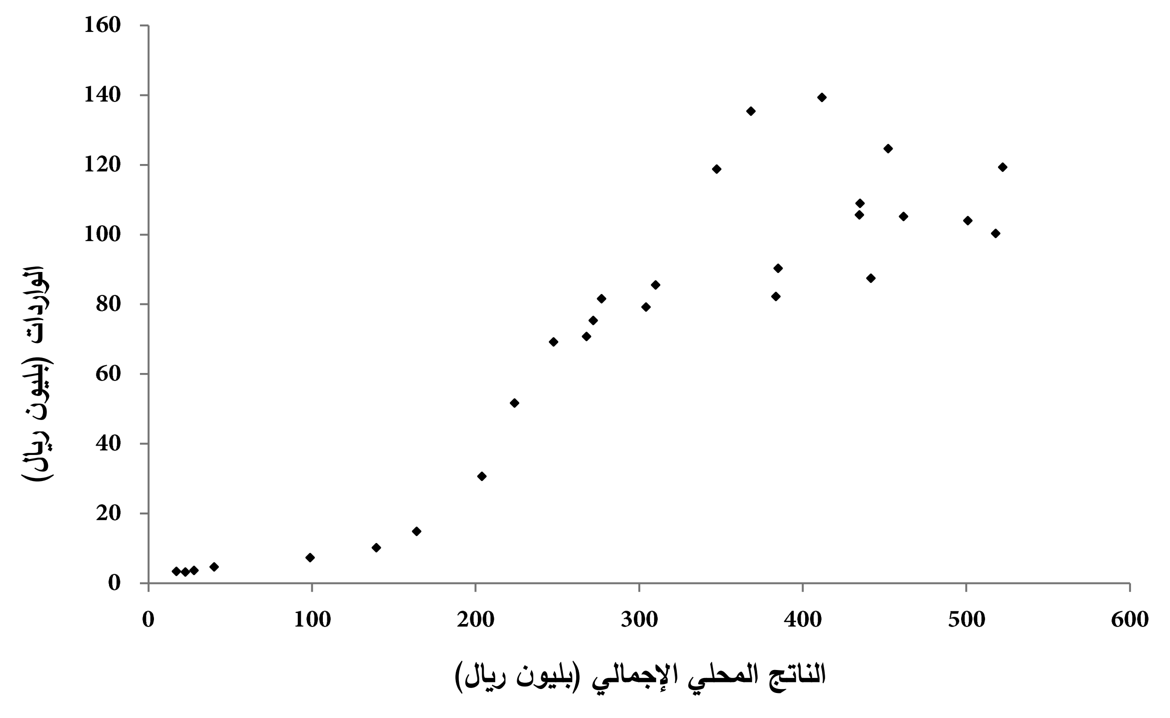
المصدر: التقرير السنوي لمؤسسة النقد العربي السعودي (العدد رقم ٣٣)، ١٩٩٧.

الحل:

يشير الشكل رقم (٧-١١) إلى وجود علاقة خطية طردية بين الواردات والناتج المحلي الإجمالي. وبإجراء انحدار الواردات (I) على الناتج المحلي الإجمالي (G) باستخدام طريقة المربعات الصغرى الاعتيادية تم التوصل إلى المعادلة التالية:

$$\begin{aligned}\hat{I}_t = \quad & -4.4704 + 0.2582\, G_t \\ & (7.4997) \quad (0.0224) \\ & (0.556) \quad (0.000) \\ & R^2 = 0.836 \\ & S = 18.5\end{aligned}$$

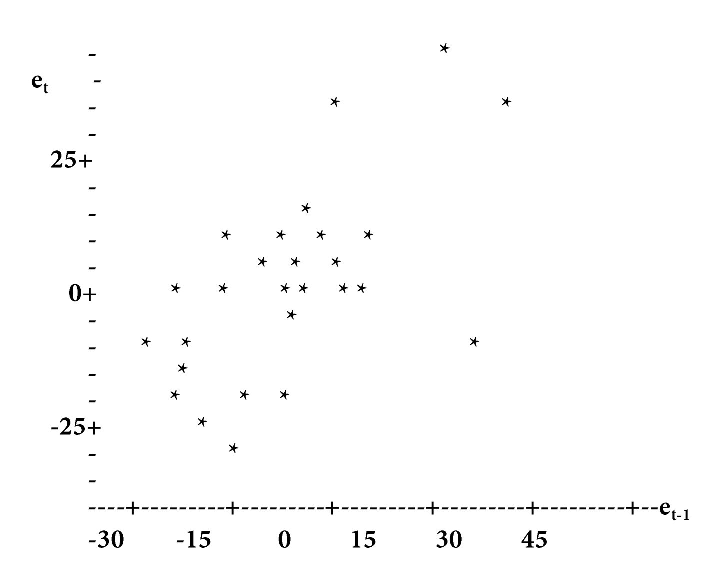
حيث إن القيم في الأقواس أسفل المقدرين هي الأخطاء المعيارية وقيم الاحتمال (p-value).



شكل رقم (٧-١١): رسم انتشار الواردات مع الناتج المحلي الإجمالي

اختبار وجود الارتباط الذاتي:

* **رسم انتشار البواقي مع البواقي المبطئة بفترة زمنية واحدة:** يوضح شكل انتشار البواقي مع البواقي المبطئة بفترة زمنية واحدة، أن هناك علاقة تربط بينهما مما يشير إلى وجود ارتباط ذاتي من الدرجة الأولى (شكل رقم (٧-١٢)).



شكل رقم (٧-١٢): رسم انتشار البواقي (e_t) مع البواقي المبطئة بفترة زمنية واحدة (e_{t-1}).

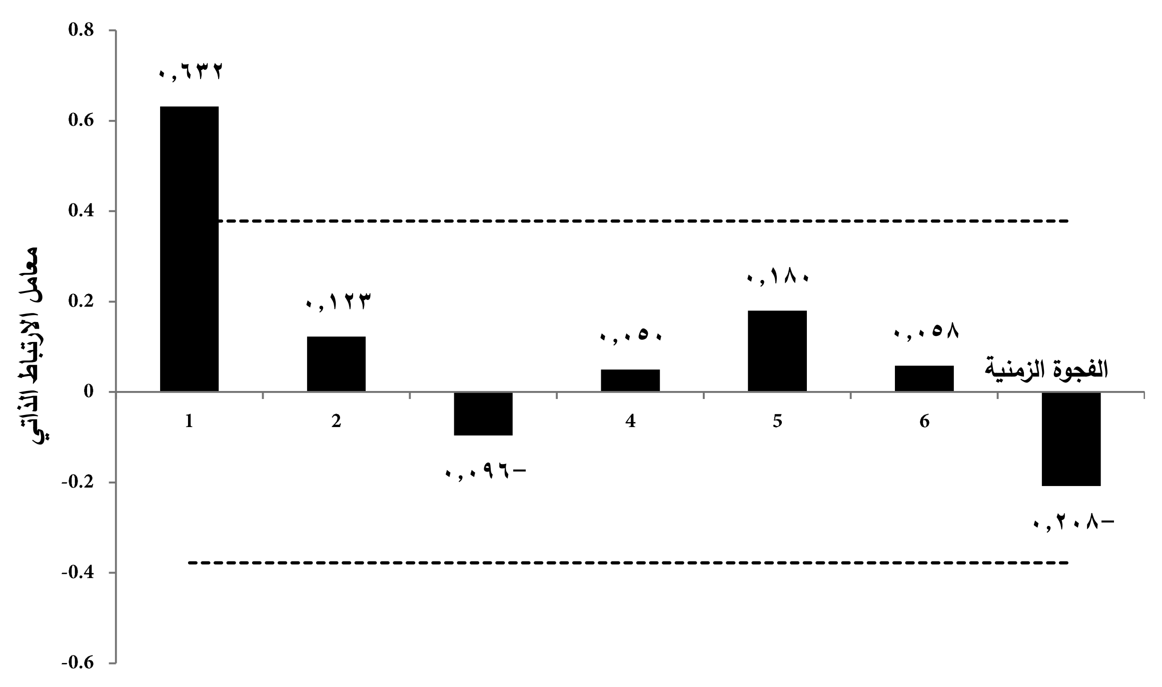
* **رسم معاملات الارتباط الذاتي مع الفجوات الزمنية:**

لرسم معاملات الارتباط مع الفجوات الزمنية، يتم حساب معامل الارتباط من الرتبة الأولى والثانية والثالثة، إلخ. فمثلاً لحساب معامل الارتباط من الرتبة الأولى AR(1) باستخدام الحسابات الموضحة بالجدول (٧-١١):

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} = \frac{5625.6936}{8900.2518} = 0.63208$$

وهكذا يتم حساب $\hat{\rho}_2$ و $\hat{\rho}_3$، ... كما يتم رسم خطين يمثلان حدي ثقة أعلى وأدنى بحساب القيمة $+\frac{2}{\sqrt{n}}$ للحد الأعلى و $-\frac{2}{\sqrt{n}}$ للحد الأدنى، أي من هذا المثال نجد أن الحد الأعلى $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{28}} = 0.38$ والحد الأدنى -٠,٣٨. ويوضح الشكل رقم (٧-١٣) رسم معاملات الارتباط الذاتي مع الفجوات الزمنية. ويلاحظ من الشكل أن معامل الارتباط الذاتي عند الفجوة ١ ($\hat{\rho}_1$) هو الوحيد الذي تختلف قيمته عن الصفر عند مستوى معنوية (٥%). أما بقية قيم معاملات

الارتباط الذاتي للفجوات التالية تأخذ قيماً سالبة وموجبة وجمعيها لا تختلف معنوياً عن الصفر عند مستوى دلالة ٥%. ويشير هذا الشكل إلى احتمال وجود ارتباط ذاتي بين بواقي نموذج الواردات من الدرجة الأولى ((AR(1)).



شكل رقم (٧-١٣): رسم معاملات الارتباط الذاتي لبواقي نموذج الواردات مع الفجوات الزمنية

- **اختبار ديربن-واتسون**: يوضح الجدول رقم (٧-١٢) القيم اللازمة لحساب إحصاء ديربن-واتسون. ومن القيم الواردة بالجدول تم حساب قيمة إحصاء (d) على النحو التالي:

$$d = \frac{\sum_{t=2}^{28}(e_t - e_{t-1})^2}{\sum_{t=1}^{28} e_t^2} = \frac{6100.212616}{8900.2518} = 0.6854$$

ومن جداول ديربن-واتسون باستخدام مستوى معنوية (١%) وعدد المشاهدات (n=28) وعدد المتغيرات المستقلة (p=1) نتحصل على القيم الحرجة التالية (انظر ملحق الجداول الإحصائية):

$$\mathbf{d_L = 1.10 \qquad d_U = 1.24}$$

وبما أن **1.10 > 0.685** نرفض فرض العدم ونقبل الفرض البديل القائل بوجود ارتباط ذاتي موجب بين بواقي النموذج.

جدول رقم (٧-١٢): القيم اللازمة لحساب إحصاء ديربن-واتسون

المشاهدة	e_t	e_{t-1}	$e_t \times e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
١	٣,٤٢١٦	-	-	-	١١,٧٠٧٢
٢	١,٨٣٩٣	٣,٤٢١٦	٦,٢٩٣٥	٢,٥٠٣٥	٣,٣٨٣٢
٣	٠,٩٤٥٨	١,٨٣٩٣	١,٧٣٩٧	٠,٧٩٨٣	٠,٨٩٤٦
٤	١,١٧٢٤-	٠,٩٤٥٨	١,١٠٨٩-	٤,٤٨٧٠	١,٣٧٤٦
٥	١٣,٧٤٤٠-	١,١٧٢٤-	١٦,١١٣٧	١٥٨,٠٤٣٥	١٨٨,٨٩٦٣
٦	٢١,٣٣٤٢-	١٣,٧٤٤٠-	٢٩٣,٢١٦٧	٥٧,٦١٢٢	٤٥٥,١٤٩٢
٧	٢٣,٠٣٢٤-	٢١,٣٣٤٢-	٤٩١,٣٧٨٤	٢,٨٨٣٨	٥٣٠,٤٩١٤
٨	١٧,٥٠٤٩-	٢٣,٠٣٢٤-	٤٠٣,١٧٩١	٣٠,٥٥٣٦	٣٠٦,٤٢٠٤
٩	١,٦٦٨٧-	١٧,٥٠٤٩-	٢٩,٢٠٩٦	٢٥٠,٧٨٥٦	٢,٧٨٤٤
١٠	٩,٧٠٥٣	١,٦٦٨٧-	١٦,١٩٤٧-	١٢٩,٣٦٥٩	٩٤,١٩٢٠
١١	١٢,٣٦٧٥-	٩,٧٠٥٣	١٢٠,٠٢٩٧-	٤٨٧,٢٠٦٢	١٥٢,٩٥٤٨
١٢	٢٨,٩٤٤٨-	١٢,٣٦٧٥-	٣٥٧,٩٧٤٦	٢٧٤,٨٠٧٣	٨٣٧,٨٠١٦
١٣	١١,٠٧٦٨-	٢٨,٩٤٤٨-	٣٢٠,٦١٦٥	٣١٩,٢٦٤٧	١٢٢,٦٩٦٠
١٤	٣٧,٤٦٧٦	١١,٠٧٦٨-	٤١٥,٠٢١٩-	٢٣٥٦,٥٦٠٠	١٤٠٣,٨٢٠٣
١٥	٤٤,٧٥٥٢	٣٧,٤٦٧٦	١٦٧٦,٨٦٧٩	٥٣,١٠٨٧	٢٠٠٣,٠٢٤٢
١٦	٣٣,٤٩٣٠	٤٤,٧٥٥٢	١٤٩٨,٩٨٥١	١٢٦,٨٣٦٠	١١٢١,٧٨١٨
١٧	٩,٩٦٨٦	٣٣,٤٩٣٠	٣٣٣,٨٧٧٣	٥٥٣,٣٩٩٦	٩٩,٣٧٢٣
١٨	٦,٠٨١١	٩,٩٦٨٦	٦٠,٦١٩٧	١٥,١١٢٥	٣٦,٩٧٩٦
١٩	٩,٥٣٩٤	٦,٠٨١١	٥٨,٠٠٩٨	١١,٩٥٩٩	٩١,٠٠٠٠
٢٠	١٤,٥٤١٤	٩,٥٣٩٤	١٣٨,٧١٦٥	٢٥,٠٢٠٥	٢١١,٤٥٣٤
٢١	٥,١٦٥١	١٤,٥٤١٤	٧٥,١٠٧٨	٨٧,٩١٦٠	٢٦,٦٧٨١
٢٢	٤,٦٦٩٠-	٥,١٦٥١	٢٤,١١٥٩-	٩٦,٧٠٩٨	٢١,٧٩٩٨
٢٣	١,٠٥٦١	٤,٦٦٩٠-	٤,٩٣١٠-	٣٢,٧٧٧٢	١,١١٥٤
٢٤	١٢,٢٧٨٩	١,٠٥٦١	١٢,٩٦٨٠	١٢٥,٩٥١٢	١٥٠,٧٧١٧
٢٥	٢,١٣٢٥-	١٢,٢٧٨٩	٢٦,١٨٤٩-	٢٠٧,٦٨٩١	٤,٥٤٧٦
٢٦	٢٢,١٥٤١-	٢,١٣٢٥-	٤٧,٢٤٣٨	٤٠٠,٨٦٣٤	٤٩٠,٨٠٣٤
٢٧	٩,٥٤٧٩-	٢٢,١٥٤١-	٢١١,٥٢٤٤	١٥٨,٩١٦٥	٩١,١٦١٩
٢٨	٢٠,٩٠٩٢-	٩,٥٤٧٩-	١٩٩,٦٣٨٨	١٢٩,٠٨٠٨	٤٣٧,١٩٦٦
المجموع			٥٦٢٥,٦٩٣٦	٦١٠٠,٢١٢٦١٦	٨٩٠٠,٢٥١٨

- **اختبار التلاحقات:** لإجراء هذا الاختبار تم أولاً حساب القيم التالية:

عدد إشارات البواقي الموجبة (N_1) = ١٤

عدد إشارات البواقي السالبة (N_2) = ١٤

عدد التلاحقات (n) = ٨

وباستخدام هذه القيم تم حساب الوسط الحسابي والتباين للتلاحقات كما يلي:

$$E(n) = \frac{2N_1N_2}{N_1+N_2} + 1 = \frac{2x14x14}{14+14}+1 = 15$$

$$\hat{\sigma}_n^2 = \frac{2N_1N_2(2N_1N_2-N_1-N_2)}{(N_1+N_2)^2(N_1+N_2-1)} = \frac{2x14x14(2x14x14-14-14)}{(14+14)^2x(14+14-1)} = 6.741$$

وبما أن الوسط الحسابي للتلاحقات يساوي (١٥) والانحراف المعياري يساوي (٢,٥٩٦)، فإن فترة ثقة (٩٥%) هي:

$$15 \pm 1.96 \times 2.596$$

أي فترة الثقة هي: ٩,٩١ – ٢٠,٠٩

وبما أن عدد التلاحقات (٨) يقع خارج هذه الفترة، فإننا نرفض فرض العدم عند مستوى دلالة ٥%، ونحكم بأن البواقي غير عشوائية.

ونخلص من نتائج هذه الاختبارات أن نموذج الطلب على الواردات يعاني من مشكلة الارتباط الذاتي. ويترتب على وجود الارتباط الذاتي كما سبق ذكره، أن تكون النتائج التي نتحصل عليها من الاستدلال الإحصائي غير صحيحة.

اختبار بروش - جودفري (Breusch-Godfrey): لإجراء الاختبار تم بناء نموذج انحدار البواقي على الناتج المحلي الإجمالي والبواقي المبطئة بفترة واحدة. ويوضح الجدول رقم (٧-١٣) شكل البيانات لنموذج البواقي. ويوضح الإطار رقم (٧-١) نتائج انحدار البواقي على الناتج المحلي الإجمالي والبواقي المبطئة. ومن نتائج النموذج تم حساب الإحصاءة:

$$n \times R^2 = 28 \times 0.4212 = 11.795$$

ويلاحظ القارئ أنه تم حساب إحصاءة الاختبار باستخدام الصيغة $n \times R^2$ بدلاً عن الصيغة $(n-q)R^2$، ذلك لأننا استخدمنا القيمة صفر للمشاهدة الأولى في متغير البواقي المبطئة بفترة واحدة، ومن ثم أصبح عدد المشاهدات (٢٨) بدلاً عن (٢٧). وحيث إن قيمة الإحصاءة (١١,٧٩٥) أكبر بكثير من القيمة الحرجة لمربع كاي عند درجة حرية واحدة ومستوى معنوية (٥%) ، $\chi^2_{0.05,1} = 3.84$ ، فإننا نرفض فرض العدم عند مستوى دلالة (٥%)، مما يعني وجود ارتباط ذاتي بين البواقي. وأما في حالة معاملة المشاهدة الأولى للبواقي المبطئة كمشاهدة مفقودة سيكون عدد المشاهدات (٢٧) ومن ثم الحصول على قيمة مختلفة لمعامل التحديد. وستصبح قيمة الإحصاءة في هذه الحالة

$\left(n-q\right)R^{2}=\left(28-1\right)\times 0.423078=11.42$ ، وهي أيضاً أكبر من القيمة الحرجة لمربع كاي عند درجة حرية واحدة ومستوى معنوية (٥%).

جدول رقم (٧-١٣): البواقي والبواقي المبطئة والناتج المحلي الإجمالي

المشاهدة	البواقي e_t	الناتج المحلي الإجمالي	البواقي المبطئة e_{t-1}
١	٣,٤٢١٦	١٧,١٥	٠,٠٠٠٠
٢	١,٨٣٩٣	٢٢,٥٨	٣,٤٢١٦
٣	٠,٩٤٥٨	٢٧,٨٦	١,٨٣٩٣
٤	١,١٧٢٤-	٤٠,٠٩	٠,٩٤٥٨
٥	١٣,٧٤٤٠-	٩٨,٨٤	١,١٧٢٤-
٦	٢١,٣٣٤٢-	١٣٩,٢٣	١٣,٧٤٤٠-
٧	٢٣,٠٣٢٤-	١٦٣,٨٩	٢١,٣٣٤٢-
٨	١٧,٥٠٤٩-	٢٠٣,٩٤	٢٣,٠٣٢٤-
٩	١,٦٦٨٧-	٢٢٣,٨٢	١٧,٥٠٤٩-
١٠	٩,٧٠٥٣	٢٤٧,٦٢	١,٦٦٨٧-
١١	١٢,٣٦٧٥-	٣٨٣,٥٩	٩,٧٠٥٣
١٢	٢٨,٩٤٤٨-	٥١٧,٩٩	١٢,٣٦٧٥-
١٣	١١,٠٧٦٨-	٥٢٢,١٨	٢٨,٩٤٤٨-
١٤	٣٧,٤٦٧٦	٤١١,٨	١١,٠٧٦٨-
١٥	٤٤,٧٥٥٢	٣٦٨,٤	٣٧,٤٦٧٦
١٦	٣٣,٤٩٣٠	٣٤٧,٤٢	٤٤,٧٥٥٢
١٧	٩,٩٦٨٦	٣١٠,٠٣	٣٣,٤٩٣٠
١٨	٦,٠٨١١	٢٦٧,٨٥	٩,٩٦٨٦
١٩	٩,٥٣٩٤	٢٧٢	٦,٠٨١١
٢٠	١٤,٥٤١٤	٢٧٦,٩١	٩,٥٣٩٤
٢١	٥,١٦٥١	٣٠٤,٠٨	١٤,٥٤١٤
٢٢	٤,٦٦٩٠-	٣٨٤,٩٩	٥,١٦٥١
٢٣	١,٠٥٦١	٤٣٥,٠٤	٤,٦٦٩٠-
٢٤	١٢,٢٧٨٩	٤٥٢,٣	١,٠٥٦١
٢٥	٢,١٣٢٥-	٤٣٤,٥٧	١٢,٢٧٨٩
٢٦	٢٢,١٥٤١-	٤٤١,٧٤	٢,١٣٢٥-
٢٧	٩,٥٤٧٩-	٤٦١,٦٢	٢٢,١٥٤١-
٢٨	٢٠,٩٠٩٢-	٥٠٠,٩٣	٩,٥٤٧٩-

إطار رقم (٧-١): نتائج انحدار البواقي على الناتج المحلي الإجمالي والبواقي المبطئة

Regression Statistics	
Multiple R	0.649025
R Square	0.421234
Adjusted R Square	0.374933
Standard Error	14.35432
Observations	28

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	3749.087	1874.544	9.097668	0.001075
Residual	25	5151.165	206.0466		
Total	27	8900.252			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-1.07344	5.823935	-0.18432	0.855253	-13.0681
et-1	0.666422	0.156232	4.265599	0.00025	0.344657
gdp	0.001947	0.017416	0.11182	0.911859	-0.03392

معالجة مشكلة الارتباط الذاتي:

لمعالجة مشكلة الارتباط الذاتي تم استخدام طريقة الفروق المعممة بحسب قيم معامل الارتباط الذاتي المقدرة بطريقة ثيل - نجار، كوكرين - أوركت، طريقة ديربن ذات المرحلتين:

طريقة ثيل - نجار (Theil-Nagar):

باستخدام المعادلة (7.72) تم حساب مقدر الارتباط الذاتي كما يلي:

$$\hat{\rho} = \frac{n^2(1-\frac{d}{2}) + (p+1)^2}{n^2-(p+1)^2} = \frac{28^2(1-0.5\times 0.6853978) + 4}{28^2-4} = 0.66580008$$

ولإجراء الانحدار تم تحويل المتغير التابع والمستقل على التوالي كما يلي:

$$I'_t = I_t - 0.66580008\, I_{t-1}$$
$$G'_t = G_t - 0.6658008\, G_{t-1}$$

كما تم حساب قيمتي الحالة الأولى للمتغيرين التابع والمستقل كما يلي:

$$I'_1 = I_1\sqrt{1-\hat{\rho}^2} = 3.38\sqrt{(1\text{-}0.66580008)^2} = 2.5219$$

و

$$G'_1 = G_1\sqrt{1-\hat{\rho}^2} = 17.15\sqrt{(1\text{-}0.66580008)^2} = 12.7961$$

وباستخدام طريقة المربعات الصغرى العادية تم الحصول على المعادلة المقدرة التالية:

$$\widehat{I}'_t = 5.2369 + 0.1920G'_t \qquad s = 13.36,\ R^2 = 0.473, \quad DW = 1.17$$

(5.0707) (0.0397)

(0.311) (0.0001)

حيث إن الأرقام بين القوسين أسفل قيم معالم النموذج المقدرة هي الأخطاء المعيارية وقيم الاحتمالات المناظرة لكل معامل. وبما أن قيمة إحصاء ديربن-واتسوان (d=1.17) تقع ما بين قيمتي الحد الأعلى والحد الأدنى الحرجة (, d_L=1.10 d_U=1.24)، لا يمكننا البت بوجود أو عدم وجود ارتباط ذاتي. ولكن يلاحظ أن قيمة إحصاء ديربن-واتسون قد ارتفعت من (0.69) إلى (1.17) دلالة على تقليل حدة الارتباط الذاتي إلى درجة مقدرة. والآن يمكننا إيجاد قيم معاملات النموذج الأصلي كما يلي:

$$\widehat{\beta}_0 = \frac{\widehat{\beta}'_0}{1-\widehat{\rho}} = \frac{5.2369}{1-0.66580008} = 15.67\ ,\ \widehat{\beta}_1 = \widehat{\beta}'_1 = 0.1920$$

وبذلك تكون المعادلة المقدرة هي:

$$\widehat{I}_t = 15.67 + 0.1920G_t$$

وكذلك يتم الحصول على الأخطاء المعيارية لمعاملي الانحدار الأصلي كما يلي:

$$s.e(\widehat{\beta}_0) = \frac{s.e(\widehat{\beta}'_0)}{1-\widehat{\rho}} = \frac{5.0707}{1-0.66580008} = 15.1725$$

$$s.e(\widehat{\beta}_1) = s.e(\widehat{\beta}'_1) = 0.0397$$

***طريقة ديربن ذات المرحلتين:**

لتقدير قيمة معامل الارتباط الذاتي تم إجراء انحدار الواردات على كل من الناتج الإجمالي والواردات لفترة سابقة (مبطئة لفترة واحدة) والناتج الإجمالي لفترة سابقة؛ وتم الحصول على النموذج المقدر التالي:

$$\hat{I}_t = -1.704 + 0.0646G_t + 0.5093I_{t-1} + 0.0706G_{t-1}; \quad s = 10.71, \; R^2 = 0.947$$

(4.7303) (0.0446) (0.1214) (0.0602)

(0.7219) (0.1610) (0.0003) (0.2527)

حيث إن الأرقام بين القوسين أسفل قيم معالم النموذج المقدرة هي الأخطاء المعيارية وقيم الاحتمالات المناظرة **لكل معامل.** ومن نتائج النموذج أعلاه نجد أن قيمة معامل الارتباط الذاتي المقدرة – معامل I_{t-1} - هي:

$$\hat{\rho} = 0.5093$$

وباستخدام قيمة معامل الارتباط المقدرة تم تحويل المتغيرات التابع والمستقلة بنفس الطريقة التي سبق شرحها. وباستخدام طريقة المربعات الصغرى الاعتيادية تم الحصول على المعادلة المقدرة التالية:

$$\hat{I}_t = \quad 4.5597 \quad + \quad 0.2265G_t \;; \quad s = 14.08 \quad R^2 = 0.653 \quad DW = 1.066$$

(11.53) (0.032)

(0.696) (0.000)

حيث إن الأرقام بين القوسين أسفل قيم معالم النموذج المقدرة هي الأخطاء المعيارية وقيم الاحتمالات المناظرة لكل معامل. ويلاحظ أنه على الرغم من التحسن الذي طرأ في قيمة إحصاء ديربن، إلا أن مشكلة الارتباط الذاتي ما زالت قائمة؛ إذ إن هذه القيمة أقل من القيمة الحرجة الدنيا (d_L=1.10).

*** طريقة كوكرين-أوركت:**

لتقدير معامل الارتباط الذاتي من الرتبة الأولى، تم توفيق نموذج بواقي نموذج الواردات على البواقي السابقة لها، وتم الحصول على النتيجة التالية:

$$\hat{e}_t = 0.66476e_{t-1}$$

وكما يمكن الحصول على تقدير قيمة معامل الارتباط الذاتي بين حدود البواقي باستخدام الصيغة التالية:

$$\hat{\rho}_1 = 1 - \frac{DW}{2} = 1 - \frac{0.6853978}{2} = 0.6573011$$

وباستخدام قيمة معامل الارتباط المقدرة (٠,٦٦٤٧) تم تحويل المتغيرات التابع والمستقلة بنفس الطريقة التي سبق شرحها. وباستخدام طريقة المربعات الصغرى الاعتيادية تم الحصول على المعادلة المقدرة التالية:

$$\hat{I}_t = 15.562 + 0.192G_t \quad ; \quad s = 13.37\,, \; R^2 = 0.474\,, \; DW = 1.17$$

(15.14) (0.040)

(0.313) (0.000)

حيث إن الأرقام بين القوسين أسفل قيم معالم النموذج المقدرة هي الأخطاء المعيارية وقيم الاحتمالات المناظرة لكل معامل. ويلاحظ على الرغم من التحسن الذي طرأ في قيمة إحصاء ديربن-واتسون، إلا أنه لا يمكن البت بوجود أو عدم وجود ارتباط ذاتي لوقوع قيمة الإحصاء في منطقة عدم اتخاذ القرار بوجود أو عدم وجود ارتباط ذاتي. ولذلك تستخدم قيم معاملات نموذج الانحدار المقدرة أعلاه في النموذج الأصلي ويتم الحصول على بواقي جديدة ومن ثم يتم حساب تقدير جديد لمعامل الارتباط الذاتي. وتستمر هذه العملية إلى الحد الذي تكون فيه الفروق بين قيم المعاملات ضئيلة. وباستخدام برنامج LIMDEP الذي يبدأ تقدير قيمة معامل الارتباط الذاتي بين حدود البواقي من قيمة ديربن-واتسون ($\hat{\rho}_1 = 1 - \frac{DW}{2} = 1 - \frac{0.6853978}{2} = 0.6573011$) تم التوصل إلى القيمة $\left(\hat{\rho}_1 = 0.878411\right)$ ، الحد الذي بدأت فيه قيمة التقدير في الاستقرار - تم التوصل إلى النموذج المقدر التالي (إطار رقم ٧-٢):

$$\hat{I}_t = 52.012 + 0.1133G_t \;; \qquad s = 12.94 \quad DW = 1.28$$

$$(30.3458) \qquad (0.0514)$$

$$(0.0865) \qquad (0.0276)$$

حيث إن الأرقام بين القوسين أسفل قيم معالم النموذج المقدرة هي الأخطاء المعيارية وقيم الاحتمالات المناظرة لكل معامل. ويلاحظ أنه الآن لا يوجد دليل وجود ارتباط ذاتي من الدرجة الأولى ((AR(1)) بين حدود الخطأ المتتالية.

إطار رقم (٧-٢): مخرجات برنامج LIMDEP لنموذج انحدار الواردات باستخدام طريقة كوكرين-أوركت

```
+-----------------------------------------------+
| AR(1) Model:     e(t) = rho * e(t-1) + u(t) |
| Initial value of rho       =        .65730 |
| Maximum iterations         =           100 |
| Method = Cochrane - Orcutt                 |
| Iter=  1, SS=   4639.216, Log-L=-111.554481 |
| Iter=  2, SS=   4275.162, Log-L=-110.624413 |
| Iter=  3, SS=   4186.812, Log-L=-110.573483 |
| Iter=  4, SS=   4231.362, Log-L=-110.909705 |
| Final value of Rho     =          .878411 |
| Iter=  4, SS=   4231.362, Log-L=-110.909705 |
| Durbin-Watson:   e(t) =          .121940 |
| Std. Deviation:  e(t) =        27.078750 |
| Std. Deviation:  u(t) =        12.941116 |
| Durbin-Watson:   u(t) =         1.281216 |
| Autocorrelation: u(t) =          .359392 |
| N[0,1] used for significance levels       |
+-----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant      52.0120080     30.3457541     1.714   .0865
 GDP             .11327688      .05142157     2.203   .0276    295.659286
 RHO             .87841060      .09197319     9.551   .0000
```

ملاحظات:

- يلاحظ أنه تم الحصول على قيم تقديرية مختلفة لمعامل الارتباط الذاتي بين البواقي باستخدام معادلة انحدار البواقي على البواقي المبطأة وباستخدام قيمة معامل ديربن-واتسون واستخدام طرق ثيل - نجار وطريقة ديربن ذات المرحلتين. وتبعاً لهذا الاختلاف كانت النتائج التي تم الحصول عليها من النماذج المصححة أيضاً مختلفة. فالسؤال، ما الطريقة المثلى لتقدير معامل الارتباط الذاتي؟ في حالة تحليل عدد كبير من المشاهدات – ما بين ٦٠ إلى ٧٠ مشاهدة - نحصل باستخدام هذه الطرق على قيم متقاربة لمعامل الارتباط (Gujarati, 1988, p.388). غير أنه في العينات الصغيرة غالباً ما نحصل على قيم تقديرية مختلفة لمعامل الارتباط الذاتي كما هو الحال في هذا المثال. ولسوء الحظ لا توجد طريقة مثلى يوصى بها في حالة العينات الصغيرة.
- من الحلول التي يمكن أن تسهم في حل مشكلة الارتباط الذاتي هو إضافة متغير/متغيرات تسهم في تفسير التغير/التباين في المتغير التابع. فمثلاً في هذا المثال نجد الطلب على الواردات دالة ليس فقط على الناتج القومي بل على حجم السكان وأسعار الواردات والأسعار القياسية للسلع ايضاً. فبإضافة مثل هذه المتغيرات فإنه يمكن التخلص من هذه المشكلة.
- **طريقة الفرق الأول (First-Difference Method):**

بما أن معامل الارتباط الذاتي تتراوح قيمته ما بين الصفر وسالب أو موجب واحد صحيح $(+1 -)$، $(-1<\rho<1)$، يمكن افتراض وجود علاقة تامة بين البواقي والبواقي المبطأة بفترة واحدة، أي أن $\hat{\rho}_1 = 1$ ، ومن ثم نحصل على المعادلة التالية للنموذج الخطي البسيط:

$$y_t - y_{t-1} = \beta_1\left(x_t - x_{t-1}\right) + \left(\varepsilon_t - \varepsilon_{t-1}\right) \quad (7\text{-}80)$$

وباستخدام مشغل الفروق (shift operator) Δ نحصل على:

$$\Delta y_t = \beta_1 \Delta x_t + u_t \quad (7\text{-}81)$$

ويلاحظ أن حد الخطأ في المعادلة (7.80) لا يحتوي على معامل الارتباط ولا يحتوي النموذج على الثابت.

وتُعد تحويلة الفرق الأول ملائمة إذا كانت قيمة إحصاء ديربن-واتسون قليلة وقيمة الارتباط الذاتي بين البواقي كبيرة. ويقترح مادالا (Madala, 2001) استخدام طريقة الفروق الأولى إذا كانت قيمة ديربن-واتسون أقل من قيمة معامل التحديد، أي $\left(d < R^2\right)$. ومن المثال نجد أن قيمة ديربن-واتسون أقل من قيمة معامل التحديد،

$$d = 0.6854 < R^2 = 0.836$$

ومن ثم يمكن إجراء انحدار الفرق الأول للمتغير التابع (الواردات) على الفرق الأول (الناتج المحلي الإجمالي)، ونحصل على المعادلة التالية:

$$\Delta\hat{y}=0.1187\Delta x \qquad R^2=0.186,\ DW=1.28$$
$$(0.0487)$$
$$(0.022)$$

حيث إن القيمتين بين الأقواس هما الخطأ المعياري وقيمة الاحتمال. ويلاحظ أن قيمة ديربن-واتسون تشير إلى عدم وجود ارتباط ذاتي بين بواقي النموذج.

- توجد طرق أخرى لتقدير معالم نموذج الانحدار في وجود ارتباط ذاتي بين حدود الخطأ برتب مختلفة (AR(1), AR(2), ...) باستخدام طريقة الإمكان الأعظم (Maximum likelihood estimation) وطريقة المربعات الصغرى غير المقيدة (Unconditional least squares) ويول-ولكر (Yule-Walker estimate). وتجدر الإشارة إلى أننا نحصل على تقديرات مختلفة لمعالم النموذج باستخدام هذه الطرق (انظر الجدول رقم ٧-١٤). كما أن هذه الطرق متوفرة في بعض برامج الإحصاء كبرنامج SAS وSHAZAM وLIMDEP.. إلخ.

جدول رقم (٧-١٤): نتائج نموذج انحدار الواردات باستخدام طرق كوكرين-أوركت والإمكان الأعظم والمربعات الصغرى غير المقيدة ويول-ولكر باستخدام نظام SAS.

المتغير	النموذج الأساسي	كوكرين-أوركت	الإمكان الأعظم	المربعات الصغرى غير المقيدة	يول-ولكر
الثابت	٤,٤٧٠٤-	٥٢,٠١٢	٢٠,١٢٦٩	٢٣,٨٥٧٣	٨,٢٨١٩
الناتج المحلي الإجمالي	٠,٢٥٨٢	٠,١١٣٣	٠,١٦٤٨	٠,١٤٧٦	٠,٢١١٢
الارتباط الذاتي AR(1)	-	٠,٨٧٨٤	٠,٨٢٢٢	٠,٨٧٩٦	-
معامل التحديد R^2	٠,٨٣٥٩	٠,١٨٩	٠,٣٧٩١	٠,٣٠٢٢	٠,٦٠٩١
معامل ديربن-واتسون	٠,٦٨٥٤	١,٢٨	١,٢٤٣٦	١,٢٤٤٦	١,١٧

٧-٥ مشكلة الارتباط الذاتي وعدم ثبات التباين في آن واحد:

تطرقنا في الجزأين (٧-٢) و(٧-٣) إلى الارتباط الذاتي وعدم ثبات التباين كمشكلتين منفصلتين عن بعض. غير أنه يمكن أن نواجه بالمشكلتين في آن واحد، أي أن يعاني النموذج الموفق من عدم ثبات التباين ووجود ارتباط ذاتي بين حدود الخطأ المتتالية. وتحدث هذه المشكلة عند تحليل بيانات سلسلة زمنية-قطاعية (Longitudinal or Panel data). مثلاً يقوم باحث باختيار عينة عشوائية من الأسر ومن ثم يتم جمع بيانات عن الدخل والمصروفات المعيشية عبر فترات زمنية متساوية (كل شهر، سنة، إلخ) وذلك لعدة شهور أو سنوات من نفس هذه الأسر.

وفي ظل مشكلتي الارتباط الذاتي وعدم ثبات التباين، يمكن تقدير معالم نموذج الانحدار الخطي وستظل مقدرات المربعات الصغرى غير متحيزة ومتسقة ولكنها تفقد خاصية الكفاءة. وتأخذ مصفوفة التغاير والتباين في حالة وجود المشكلتين معاً الصيغة التالية:

$$E(\varepsilon\varepsilon^{T})=\begin{pmatrix}\sigma_1^2 & E(\varepsilon_1\varepsilon_2) & E(\varepsilon_1\varepsilon_3) & ... & E(\varepsilon_1\varepsilon_N)\\ E(\varepsilon_2\varepsilon_1) & \sigma_2^2 & E(\varepsilon_2\varepsilon_3) & .... & E(\varepsilon_2\varepsilon_N)\\ . & . & . & . & .\\ . & . & . & . & .\\ E(\varepsilon_N\varepsilon_1) & E(\varepsilon_N\varepsilon_2) & E(\varepsilon_N\varepsilon_3) & ... & \sigma_N^2\end{pmatrix} \quad \textbf{(7-82)}$$

حيث إن:

$$E(\varepsilon_i^2)=\sigma_i^2 \qquad i=1,2,...,N$$
$$E(\varepsilon_i\varepsilon_j)\neq 0 \text{ for all } i,j=1,2,...,N$$

وللكشف عن وجود المشكلتين تستخدم نفس طرق الكشف لكل مشكلة على حدة كما سبق شرحهما في الجزأين (٧-٢-٤) و(٧-٣-٥).

طريقة المربعات الصغرى المعممة لحل مشكلتي عدم ثبات التباين والارتباط الذاتي:

لصياغة نموذج انحدار لبيانات سلسلة زمنية-قطاعية، نفترض أن لدينا عدد g وحدة أو مجموعة تأخذ القيم (i=1,2,…,g) وm فترة زمنية متتالية (t=1,2,…,m)، أي أن مجموع المشاهدات يساوي (n=gm). وعليه يمكن ترميز المتغيرات كما يلي:

Y_{it} = قيمة المتغير التابع للمجموعة i في الفترة t.

X_{jit}= قيمة المتغير المستقل رقم j للمجموعة i في الفترة t.

ويأخذ نموذج الانحدار الصيغة التالية:

$$y_{it}=\beta_0+\beta_1x_{1it}+\beta_2x_{2it}+...+\beta_px_{pit}+\varepsilon_{it} \qquad (7-83)$$

وباستخدام صيغ المصفوفات يمكن تنظيم المعادلة (7.75) بحسب المجموعات كما يلي:

$$\mathbf{y}_i=\begin{bmatrix}y_{i1}\\ .\\ .\\ y_{im}\end{bmatrix},\ \mathbf{x}_i=\begin{bmatrix}x_{1i1} & x_{2i1} & ... & x_{pi1}\\ . & . & ... & .\\ x_{1im} & x_{2im} & ... & x_{pim}\end{bmatrix},\ \boldsymbol{\varepsilon}_i=\begin{bmatrix}\varepsilon_{i1}\\ .\\ .\\ .\\ \varepsilon_{im}\end{bmatrix}$$

كما يمكن تجميع البيانات بحسب المجموعات كما يلي:

$$\mathbf{y}=\begin{bmatrix} y_1 \\ . \\ . \\ . \\ y_g \end{bmatrix},\ \mathbf{x}=\begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_g \end{bmatrix},\ \boldsymbol{\varepsilon}=\begin{bmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_g \end{bmatrix}$$

حيث إن:

$\mathbf{y}$ = متجه عمودي من الدرجة nx1 يحتوي على قيم المتغير التابع.

$\mathbf{x}$ = مصفوفة البيانات من الدرجة nxp

$\boldsymbol{\varepsilon}$ = متجه عمودي من الدرجة nx1 يحتوي على حد الخطأ العشوائي.

والآن يمكن كتابة المعادلة (7.75) كما يلي:

$$\mathbf{y}=\begin{bmatrix}\mathbf{I} & \mathbf{x}\end{bmatrix}\begin{bmatrix}\beta_0 \\ \boldsymbol{\beta}\end{bmatrix}+\boldsymbol{\varepsilon}$$

حيث إن:

$\mathbf{I}$= متجه عمودي من الدرجة nx1 يحتوي عناصره من الواحد الصحيح.

$\mathbf{x}$= مصفوفة البيانات من الدرجة nxp.

β_0= المعامل الثابت.

$\boldsymbol{\beta}$= بقية المعالم $(\beta_1\ \beta_2\ \dots\ \beta_p)^T$.

وفي حالة عدم ثبات التباين ووجود ارتباط ذاتي تكون القيم المتوقعة لحدود الخطأ كما يلي:

$$E(\varepsilon_{it}^2)=\sigma_{ii} \qquad \text{for all t; i=1,2,...,g}$$
$$E(\varepsilon_{it}\varepsilon_{jt})=\sigma_{ij} \qquad \text{for all t and } i \neq j$$
$$E(\varepsilon_{it}\varepsilon_{js})=0 \qquad \text{for all i,j , and } t \neq s$$

وبالتالي تأخذ مصفوفة التغاير والتباين الشكل التالي:

$$E(\varepsilon\varepsilon^{T}) = \mathbf{V} = \begin{pmatrix} \sigma_{11}\mathbf{I}_m & \sigma_{12}\mathbf{I}_m & \cdots & \sigma_{1g}\mathbf{I}_m \\ \sigma_{21}\mathbf{I}_m & \sigma_{22}\mathbf{I}_m & \cdots & \sigma_{2g}\mathbf{I}_m \\ . & . & \cdots & . \\ \sigma_{1g}\mathbf{I}_m & \sigma_{2g}\mathbf{I}_m & \cdots & \sigma_{gg}\mathbf{I}_m \end{pmatrix} \qquad \textbf{(7-84)}$$

والآن يمكننا استخدام طريقة المربعات الصغرى المعممة لتقدير معالم النموذج (7.83) كما يلي:

$$\hat{\beta}^* = \left(\mathbf{x}^T\mathbf{v}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{v}^{-1}\mathbf{y}$$

وبما أن قيم σ_{ij} غير معلومة، فيتم تقديرها باستخدام المعادلة التالية:

$$s_{ij} = \frac{e_i^T e_j}{m\text{-}p\text{-}1}$$

حيث إن e_i الجزء رقم i من متجه بواقي النموذج (7.83) المقدر باستخدام طريقة المربعات الصغرى الاعتيادية. وبتعويض قيم S_{ij} في المصفوفة (7.84) نحصل على المصفوفة المقدرة $\hat{\mathbf{V}}$ ومن ثم يتم حساب مقدرات المربعات الصغرى المعممة.

تمارين

١. يرغب أحد الباحثين في معرفة ما إذا كان دخل الأسرة وعدد الأطفال وعدد أفراد الأسرة يؤثر على الإنفاق على ترفيه وتسلية الطفل. يعرض الجدول التالي بيانات عن هذه المتغيرات لعدد ٢٥ تلميذاً تم اختيارهم عشوائياً من إحدى المدارس الابتدائية:

مسلسل	الإنفاق على الترفيه (ريال سعودي)	عدد أفراد الأسرة	عدد الأطفال	دخل الأسرة (ريال سعودي)
1	900	5	2	10000
2	700	5	2	8500
3	600	10	3	7000
4	600	9	3	7500
5	450	11	4	6500
6	350	7	4	5500
7	200	8	4	4500
8	250	8	4	4700
9	150	8	5	4200
10	175	5	2	6000
11	180	8	3	4500
12	210	8	3	6000
13	210	10	4	6100
14	310	8	3	6400
15	410	11	3	7000
16	325	8	2	5100
17	280	7	3	4100
18	235	11	4	4000
19	190	7	5	3500
20	145	6	5	3500
21	100	12	5	3200
22	55	10	4	3000
23	55	15	6	3000
24	45	15	6	3000
25	120	10	5	3500

المطلوب:

- تقدير معادلة انحدار الإنفاق على ترفيه وتسلية الطفل (Y) على عدد الأطفال (X_1) وعدد أفراد الأسرة (X_2) ودخل الأسرة (X_3).
- هل يعاني النموذج الذي تم توفيقه من مشكلة الارتباط الخطي المتعدد؟
- إذا تبين أن النموذج يعاني من مشكلة الارتباط الخطي المتعدد، اقترح حلاً مناسباً.

٢. استخدم طريقة (Hildreth-Lu procedure) لتقدير معامل الارتباط الذاتي بين حدود الخطأ لنموذج انحدار الواردات على الناتج المحلي الاجمالي الوارد بالمثال (٧-٤-٧) ومن ثم أجرِ حل نموذج الانحدار المصحح وقارن الحل بالحلول الموضحة بالمثال.

٣. في طريقة كوكرين-أوركت، يتم تقدير معامل الارتباط الذاتي بإجراء انحدار البواقي على البواقي المبطئة بفترة زمنية واحدة، أي أن:

$$e_t=\hat{\rho}e_{t-1}+v_t$$

برهن على أنه يمكن الحصول على نفس قيمة معامل الارتباط الذاتي المقدرة باستخدام المعادلة التالية:

$$\hat{\rho}=\frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=2}^{n} e_{t-1}^2}$$



الفصل الثامن

تأكيد صحة نموذج الانحدار الخطي وعرض نتائجه

يتناول هذا الفصل آخر موضوعين في النمذجة الإحصائية بشكل عام، هما: تأكيد صحة النموذج الذي تم بناؤه وكيفية عرض نتائجه. فبعد بناء النموذج واستيفائه لجميع اشتراطات نموذج الانحدار الخطي، يتم عادةً التأكد من صحة النموذج كخطوة أخيرة تمهيداً لاستخدامه والاستفادة من نتائجه.

٨-١ تأكيد صحة النموذج:

مرحلة تأكيد صحة النموذج (Model validation) هي آخر وأهم مراحل بناء نموذج الانحدار الخطي التي تهدف إلى التأكد من مدى استقرار تقديرات معالم النموذج والاطمئنان على صحته والنتائج المترتبة على تطبيقه مقارنة بالنظرية أو الدراسات والبحوث السابقة. ويجب الإشارة إلى أن هناك اختلافاً بين تأكيد صحة النموذج وفحص النموذج (Model checking). ففحص النموذج يهدف إلى التأكد من مدى استيفاء النموذج الذي تم بناؤه لاشتراطات نموذج الانحدار، مثل التأكد من عدم وجود المشكلات القياسية كخطأ توصيف النموذج، والارتباط الخطي المتعدد، والارتباط الذاتي واختلاف التباين،.. إلخ. في حين يهدف تأكيد صحة النموذج إلى سلامة تطبيق النموذج في البيئة المستهدفة سواء كان الهدف هو التقدير أو التنبؤ أو التحكم.

توجد ثلاثة إجراءات تستخدم للتأكد من صحة النموذج، هي: جمع بيانات جديدة بهدف التأكد من قدرة أداء النموذج التنبؤية، ومقارنة نتائج النموذج الموفق بالنظريات ونتائج البحوث والدراسات السابقة ذات العلاقة ونتائج المحاكاة، وتقسيم البيانات (Data splitting) بغرض التأكد من قدرة أداء النموذج التنبؤية.

٨-١-١ جمع بيانات جديدة:

يُعد جمع بيانات جديدة من أفضل الطرق لتأكيد صحة نموذج الانحدار الموفق. وتهدف هذه الطريقة للتأكد من أن النموذج الذي تم بناؤه من بيانات سابقة ما زال صالحاً لتطبيقه على بيانات جديدة. فإذا أمكن الحصول على تنبؤات دقيقة للبيانات الجديدة، فإن استخدام النموذج الموفق يعزز ثقة المستخدم في تطبيقه. ويقترح مونتجومري وبيك وفيننج (Montgomery, Peck and Vining, 2001) أن يتم جمع ما بين (١٥) إلى (٢٠) مشاهدة جديدة لقياس مدى قدرة النموذج التنبؤية. وتوجد ثلاث طرق لاستخدام البيانات الجديدة للتأكد من صحة النموذج، هي كالتالي:

أ- من الطرق المستخدمة في تأكيد الصلاحية أن يتم إعادة تقدير معالم نموذج الانحدار باستخدام نفس الصيغة الدالية للنموذج الموفق باستخدام البيانات الجديدة. ومن ثم مقارنة معاملات الانحدار والإحصاءات الأخرى التي تم الحصول عليها من البيانات الجديدة بتلك التي تم الحصول عليها من البيانات الأصلية. فإذا كانت النتائج متقاربة، يمكن أن نستنتج أن النموذج الموفق يمكن تطبيقه في نطاق أوسع من نطاق المشاهدات الأصلية.

ب- بناء عدة نماذج انحدار باستخدام البيانات الجديدة ومن ثم مقارنتها بالنموذج الذي تم بناؤه من البيانات القديمة، فإذا كانت النماذج الجديدة تتضمن نموذجاً مماثلاً للنموذج الذي تم توفيقه من البيانات القديمة كان ذلك دلالةً على صلاحية النموذج.

ج- باستخدام البيانات الجديدة يتم حساب متوسط مربع خطأ التنبؤ (Mean squared prediction error) أو اختصارا (MSPE) باستخدام الصيغة التالية:

$$\text{MSPE} = \frac{\sum_{i=1}^{n_2}\left(y_i - \hat{y}_i\right)^2}{n_2} \qquad (8\text{-}1)$$

حيث إن:

n_2 عدد مشاهدات البيانات الجديدة.

y_i القيم الحقيقية للمتغير التابع رقم i من البيانات الجديدة.

$\hat{y}_i$ القيم المقدرة للمتغير التابع رقم i من البيانات الجديدة باستخدام نموذج الانحدار المقدر من البيانات الأصلية.

فإذا كانت قيمة متوسط مربع خطأ التنبؤ قريبة من قيمة متوسط مربعات الخطأ (MSE) في النموذج الأصلي دل ذلك على صلاحية النموذج وملاءمة استخدامه وتطبيقه.

٨-١-٢ مقارنة نتائج النموذج الموفق بالنظريات ونتائج البحوث والدراسات ذات العلاقة ونتائج المحاكاة:

للتأكد من صحة النموذج يُنصح بدراسة معاملات النموذج المقدر لتحديد مدى استقرار المعاملات من خلال تقييم إشارات وقيم المعاملات. حيث يمكن الاستفادة من الخبرة السابقة والنظريات في التأكد من صحة إشارات وحجم قيم المعاملات. فإشارة معامل النموذج غير المتوقعة أو كبر حجم القيمة المطلقة لمعامل الانحدار تشير إلى عدم ملاءمة النموذج وربما تكون نتيجة لمشكلات سوء التوصيف أو وجود مشكلة الارتباط الخطي المتعدد (Multicollinearity). كما أن القيم المتنبأ بها من المقاييس المهمة التي تشير إلى مدى ملاءمة النموذج الذي تم بناؤه. فإذا كانت القيم المتنبأ سالبة في حين أن قيم المشاهدات التي تم استخدامها في بناء النموذج موجبة، كمتغير الوزن مثلاً، فيشير ذلك إلى عدم ملاءمة تطبيق النموذج مما يستدعي الرجوع إلى مرحلة فحص النموذج للتأكد من استيفاء اشتراطات نموذج الانحدار الخطي.

٨-١-٣ تقسيم البيانات:

على الرغم من أن طريقة جمع بيانات جديدة لتأكيد صلاحية النموذج هي الطريقة الأفضل لتأكيد صحة النموذج، إلا أنها قد تكون غير عملية في أحيان كثيرة. فجمع بيانات جديدة يتطلب وقتاً وتكلفة إضافيين فضلاً عن صعوبة إجراء مسح إضافي للحصول على بيانات جديدة في بعض الحالات. ففي حالة تعذر جمع بيانات جديدة، توجد طريقة أخرى بديلة تعرف بتقسيم البيانات (Data splitting). وفي طريقة تقسيم البيانات يتم تقسيم مشاهدات البيانات الأصلية التي استخدمت في بناء نموذج الانحدار إلى مجموعتين*، الأولى تسمى مجموعة بناء النموذج (Model-building set) وتسمى المجموعة الثانية بمجموعة تأكيد صحة النموذج (Validation set). وفي الخطوة الثانية يتم بناء نموذج الانحدار لمشاهدات مجموعة بناء النموذج ومن ثم تستخدم بيانات مجموعة تأكيد صحة النموذج في التنبؤ ومن ثم يتم حساب متوسط مربع خطأ التنبؤ (MSPE). فإذا كانت قيمة متوسط مربع خطأ التنبؤ قريبة من متوسط مربع البواقي (الخطأ) لنموذج مجموعة بناء النموذج، كان ذلك دلالة على تأكيد صحة النموذج الذي تم بناؤه من البيانات الأصلية.

ملاحظات:

١. يعاب على طريقة تقسيم البيانات أنها تخفض حجم عينة مجموعة بناء النموذج مما قد يؤثر في دقة النتائج خاصة في حالة صغر حجم عينة البيانات الأصلية، وخاصة إذا كان عدد مشاهدات مجموعة بناء النموذج بعد التقسيم أصبح أقل من خمسة أضعاف عدد المتغيرات المستقلة.

٢. في حالة البيانات المقطعيِة يتم تقسيم البيانات وفق خاصية معينة للبيانات وفي حالة بيانات السلاسل الزمنية يتم تقسيم البيانات بناء على الزمن على فترتين، تمثل الفترة الأولى مجموعة بناء النموذج وتمثل الفترة الثانية مجموعة تأكيد صحة البيانات.

مثال:

من مثال بيانات انحدار وزن الطفل على عمره (الجدول رقم ٢-٢)، تم استخدام طريقة تقسيم البيانات لتأكيد صحة النموذج الذي تم بناؤه في الفصل الثاني. وفيما يلي نتائج النموذج الذي تم بناؤه في الفصل الثاني من (٥٠) مشاهدة (إطار رقم ٨-١):

وباستخدام برنامج SPSS تم اختيار عينة عشوائية قوامها (٣٠) مشاهدة لبناء النموذج، وبقية المشاهدات وعددها (٢٠) مشاهدة استخدمت لتأكيد صحة النموذج (الجدولان رقم ٨-١ و٨-٢).

ويوضح الإطار رقم (٨-٢) نتائج نموذج انحدار وزن الطفل على عمره من مشاهدات مجموعة بناء النموذج. ويتضح من نتائج نموذج الانحدار أن جميع مقدرات معالم انحدار مجموعة بناء النموذج مقاربة لمقدرات معالم النموذج الأصلي. وتوضح النتائج أن قيم معامل التحديد ومقدر التباين ومعاملي الانحدار (الثابت والعمر) في النموذجين الأصلي والنموذج الذي تم بناؤه من مجموعة بناء النموذج قريبة جداً.

* يمكن أن يتم التقسيم وفق خاصية محددة للبيانات (انظر Montgomery et al., 2001, p.535) أو بصورة عشوائية (Stevens,2012, p.93)

إطار رقم (٨-١): نتائج نموذج انحدار وزن الطفل على عمره (عدد المشاهدات=٥٠)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.962[a]	.926	.924	1.52462020

a. Predictors: (Constant), Age (years)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1386.625	1	1386.625	596.535	.000[b]
	Residual	111.574	48	2.324		
	Total	1498.199	49			

a. Dependent Variable: weight (kg)

b. Predictors: (Constant), Age (years)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.210	.325		12.948	.000
	Age (years)	2.532	.104	.962	24.424	.000

a. Dependent Variable: Weight (kg)

مصدر البيانات: مستشفى أبها للنساء والولادة والمراكز الصحية (١٩٩٩). بيانات عن أوزان ٥٠ طفلاً غير منشورة

جدول رقم (٨-١):
أوزان وأعمار (٣٠) طفلاً من منطقة عسير (مجموعة بناء النموذج)

مسلسل	الوزن	العمر	مسلسل	الوزن	العمر
١	١١,٥	٣,٠٠	٢٢	٢١,٠	٦,١٧
٣	٦,٥	٠,٥٠	٢٣	١٢,٠	٣,٠٠
٤	١٧,٠	٤,٠٠	٢٤	١٧,٥	٥,٢٥
٥	٨,٥	١,٣٣	٢٩	٤,٥	٠,٢٥
٦	٨,٨	١,٠٠	٣٠	١٥,٥	٤,٧٥
٧	٢٢,٠	٦,١٧	٣١	١٦,٥	٤,٦٧
٩	١٢,٥	٣,٦٧	٣٤	١٤,٦	٤,٨٣
١٠	١٥,٥	٥,٤٢	٣٥	١٠,٠	٢,٠٠
١١	٩,٥	١,١٧	٣٨	٨,٠	١,٠٠
١٥	١٩,٠	٦,٢٥	٤٠	١٤,٠	٣,٧٥
١٦	٩,٠	١,٥٠	٤٢	٣,٢	٠,٠٨
١٧	١٤,٠	٤,٢٥	٤٣	٥,٦	٠,٣٣
١٨	١٠,٥	٢,٠٠	٤٧	٤,٥	٠,٠٨
١٩	٦,٠	٠,٤٢	٤٨	٣,٣	٠,٠٢
٢٠	١٥,٠	٥,٥٨	٥٠	١,٤	٠,٠٨

جدول رقم (٨-٢):

أوزان وأعمار (٢٠) طفلاً من منطقة عسير (مجموعة تأكيد صحة النموذج)

مسلسل	الوزن	العمر	مسلسل	الوزن	العمر
٢	١٦,٠	٥,٠٠	٣٢	١١,٠	١,٧٥
٨	١٣,٠	٣,٤٢	٣٣	١٧,٥	٥,٢٥
١٢	١٥,٥	٤,٤٢	٣٦	٤,٠	٠,١٧
١٣	٩,٥	١,١٧	٣٧	٣,٥	٠,٠٨
١٤	١٤,٥	٢,٧٥	٣٩	٨,٠	١,٣٣
٢١	١٣,٠	٣,٤٢	٤١	١,٨	٠,١٧
٢٥	٥,٥	٠,٣٣	٤٤	٢,٨	٠,٠٨
٢٦	٥,٣	٠,٣٣	٤٥	١,٤	٠,٠١
٢٧	٦,٥	٠,٧٥	٤٦	٥,٥	٠,٥٨
٢٨	١٣,٥	٣,٨٣	٤٩	٣,٣	٠,٠٠

إطار رقم (٨-٢):

نتائج نموذج انحدار وزن الطفل على عمره من مشاهدات مجموعة بناء النموذج (عدد المشاهدات=٣٠)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.963[a]	.927	.924	1.51901146

a. Predictors: (Constant), Age (years)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	818.082	1	818.082	354.548	.000[b]
	Residual	64.607	28	2.307		
	Total	882.689	29			

a. Dependent Variable: : Weight (kg)

b. Predictors: (Constant), Age (years)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.566	.449		10.160	.000
	Age (years)	2.421	.129	.963	18.829	.000

a. Dependent Variable: Weight (kg)

ويوضح الجدول رقم (٨-٣) الحسابات اللازمة لحساب متوسط مجموع مربعات خطأ التنبؤ. وباستخدام المعادلة (8.1) نحصل على التالي:

$$\text{MSPE} = \frac{\sum_{i=1}^{20}\left(y_i - \hat{y}_i\right)^2}{20} = \frac{50.101}{20} = 2.505$$

ويلاحظ أن قيمة متوسط مجموع مربعات خطأ التنبؤ قريبة جداً من مقدر التباين (متوسط مجموع مربعات الخطأ/البواقي) في النموذج الأصلي الذي يبلغ (٢,٣٢٤). ومن هذه النتيجة نستنتج أن النموذج الأصلي الذي تم بناؤه من (٥٠) مشاهدة يمكن الاعتماد عليه واستخدامه في تقدير وزن الطفل باستخدام متغير العمر في المدى من الميلاد إلى عمر (٦,٢٥) سنة.

جدول رقم (٨-٣):
تفاصيل حساب مجموع مربعات خطأ التنبؤ (عدد المشاهدات=٢٠)

الوزن (كجم) y_i	العمر (كجم) x_i	$\hat{y}_i = 4.566 + 2.421x_i$	$\left(\hat{y}_i - y_i\right)^2$
١٦,٠	٥,٠٠	١٦,٦٧٠	٠,٤٤٩
١٣,٠	٣,٤٢	١٢,٨٤٥	٠,٠٢٤
١٥,٥	٤,٤٢	١٥,٢٦٦	٠,٠٥٥
٩,٥	١,١٧	٧,٣٩٨	٤,٤١٧
١٤,٥	٢,٧٥	١١,٢٢٣	١٠,٧٣٦
١٣,٠	٣,٤٢	١٢,٨٤٥	٠,٠٢٤
٥,٥	٠,٣٣	٥,٣٦٥	٠,٠١٨
٥,٣	٠,٣٣	٥,٣٦٥	٠,٠٠٤
٦,٥	٠,٧٥	٦,٣٨٢	٠,٠١٤
١٣,٥	٣,٨٣	١٣,٨٣٨	٠,١١٤
١١,٠	١,٧٥	٨,٨٠٣	٤,٨٢٩
١٧,٥	٥,٢٥	١٧,٢٧٦	٠,٠٥٠
٤,٠	٠,١٧	٤,٩٧٨	٠,٩٥٦
٣,٥	٠,٠٨	٤,٧٦٠	١,٥٨٧
٨,٠	١,٣٣	٧,٧٨٦	٠,٠٤٦
١,٨	٠,١٧	٤,٩٧٨	١٠,٤١٧
٢,٨	٠,٠٨	٤,٧٦٠	٤,٠٣٩
١,٤	٠,٠١	٤,٥٩٠	١٠,٤٩٩
٥,٥	٠,٥٨	٥,٩٧٠	٠,٢٢١
٣,٣	٠,٠٠	٤,٥٦٦	١,٦٠٣
	المجموع		٥٠,١٠١

٨-٢ عرض نتائج نموذج الانحدار الخطي:

يُعد عرض نتائج التحليل الإحصائي من الموضوعات المهمة التي يجب أن تُولى اهتماماً كبيراً من قبل الباحثين. فمع توفر حزم البرامج الإحصائية المختلفة، يحصل المحلل الإحصائي على مخرجات بطرق عرض مختلفة. فبالإضافة إلى طرق العرض المختلفة لمخرجات التحليل الإحصائي توفر بعض البرامج الجاهزة مخرجات كثيرة قليلة الفائدة والاستخدام. حيث درج بعض الباحثين على عرض نتائج التحليل الإحصائي كما يحصل عليها دون تعديل. لذا يُلاحظ وجود اختلاف كبير في عرض النتائج في المقالات والبحوث المنشورة في الدوريات العلمية المختلفة. ولضرورة عرض المخرجات المهمة وتوحيد عناصرها، تشترط بعض الدوريات والجمعيات العلمية أن يكون عرض نتائج التحليل الإحصائي على نسق محدد.

وفيما يلي الطرق المثلى لعرض نتائج نموذج الانحدار الخطي باستخدام مثال تحليل انحدار وزن الطفل على العمر والوزن (الفصل الثالث).

٨-٢-١ عرض نتائج نموذج الانحدار الخطي في جدول:

- طريقة جمعية علم النفس الأمريكية:

يتم عرض نموذج الانحدار عادة في شكل جدول. ويعد نظام جمعية علم النفس الأمريكية (American Psychological Association (APA في عرض نتائج نموذج الانحدار من أهم الطرق المتبعة في كثير من الدوريات الرصينة. ويوضح الإطار رقم (٨-٣) طريقة عرض جمعية علم النفس الأمريكية لنتائج نموذج الانحدار (Smith, Gratz, & Bousquet, 2009). ويلاحظ من طريقة العرض أنه يتم استخدام منزلتين عشريتين فقط ويشار إلى قيمة الاحتمال بنجوم (ثلاثة نجوم إذا كانت قيمة الاحتمال أقل من ٠,٠٠١ ونجمتين إذا كانت القيمة أقل من ٠,٠١ ونجمة واحدة إذا كانت القيمة أقل من ٠,٠٥)؛ ويشار إلى عدد المشاهدات المستخدمة في التحليل في عنوان الجدول، كما لا يتم عرض الثابت (constant) والإحصاءات المرتبطة به. وحسب نظام الجمعية لا تستخدم خطوط عمودية في الجدول. ويوضح الإطار رقم (٨-٤) عرض نموذج الانحدار حالة وجود متغيرات غير دالة إحصائياً.

إطار رقم (٨-٣): عرض نتائج نموذج الانحدار الخطي وفق نظام جمعية علم النفس الأمريكية
نموذج انحدار وزن الطفل على طوله وعمره (عدد المشاهدات =٥٠)

المتغير	المعامل	الخطأ المعياري	معامل الانحدار المعياري
الطول (سم)	٠,١٢	٠,٠٢	٠,٥٤***
العمر (سنة)	١,٢٠	٠,٢١	٠,٤٦***

معامل التحديد R^2 = ٠,٩٦؛ معامل التحديد المعدل $\bar{R}^2$=٠,٩٦

*** قيمة الاحتمال أقل من ٠,٠٠١ $(p < .001)$

إطار رقم (٨-٤): عرض نتائج نموذج الانحدار الخطي وفق نظام جمعية علم النفس الأمريكية، حالة عدم وجود متغيرات غير دالة إحصائياً

جدول X

نموذج انحدار المصروفات المعيشية على مستوى تعليم رب الأسرة وعدد الأطفال ودخل الأسرة وعدد أفراد الأسرة (عدد المشاهدات =٣٠)

المتغير	المعامل	الخطأ المعياري	معامل الانحدار المعياري
مستوى تعليم رب الأسرة (عدد السنوات الدراسية)	٠,٠٩	٠,٠٧	٠,١٣
عدد الأطفال	٠,٤٥	٠,٢٣	٠,٢٤
دخل الأسرة (ألف ريال)	٠,٤٤	٠,٠٨	٠,٦٩***
عدد أفراد الأسرة	-٠,٠٦	٠,٢٢	-٠,٠٤

معامل التحديد R^2 = ٠,٨٠؛ معامل التحديد المعدل $\bar{R}^2$=٠,٧٧

*** قيمة الاحتمال أقل من ٠,٠٠١ $(p < .001)$

- طرق جدولية أخرى:

يوضح الإطار رقم (٨-٥) طريقة لعرض نتائج نموذج الانحدار الخطي في شكل جدول اقترحها كل من تاباشنك وفيدل (Tabachnick and Fidell, 2007) وهي التي تناسب النشر في الدوريات العلمية.

إطار رقم (٨-٥): عرض نتائج نموذج الانحدار الخطي (Tabachnick and Fidell, 2007)

نموذج انحدار وزن الطفل على طوله وعمره

المتغير	الوزن (كجم)	الطول (سم)	العمر (سنة)	معامل الانحدار	المعامل المعياري	مربع معامل الارتباط شبه الجزئي Semipartial correlation
الطول	٠,٩٧	-	-	٠,١٢٥	٠,٥٤١***	٠,٠٣٧
العمر	٠,٩٦	٠,٩٤	-	١,٢٠١	٠,٤٥٦***	٠,٠٢٦
الثابت	-	-	-	-٢,١٨٢	-	-
المتوسط	١٠,١٥	٧٦,٤٠	٢,٣٥			
الانحراف المعياري	٥,٥٣	٢٤,٠١	٢,١٠			

معامل التحديد R^2 = ٠,٩٦

معامل التحديد المعدل $\bar{R}^2$=٠,٩٦

معامل الارتباط المتعدد R = ٠,٩٨***

*** قيمة الاحتمال أقل من ٠,٠٠١ $(p < .001)$

ويتضح من الإطار، أن عرض تاباشنك وفيدل لنتائج نموذج الانحدار يشمل معلومات أكثر، تشمل الإحصاء الوصفي (المتوسط والانحراف المعياري) وقيم معامل الارتباط الخطي البسيط ومربع معامل الارتباط شبه الجزئي (Semipartial correlation).

يتم عرض نتائج نموذج الانحدار الخطي وفقاً لشاترجي وهادي (Chatterjee and Hadi, 2006) كما في الإطار رقم (٨-٦). ويلاحظ أن الإطار يحتوي على معلومات أكثر لنتائج نموذج الانحدار من عرض جمعية علم النفس الأمريكية.

إطار رقم (٨-٦): عرض نتائج نموذج الانحدار (Chatterjee & Hadi, 2006)

نموذج انحدار وزن الطفل على طوله وعمره (عدد المشاهدات =٥٠)

المتغير	المعامل	الخطأ المعياري	قيمة t	قيمة الاحتمال
الثابت	٢,١٨٢-	٠,٩٧٦	٢,٢٣٥-	٠,٠٣٠
الطول (سم)	٠,١٢٥	٠,٠١٨	٦,٧٤٥	٠,٠٠٠
العمر (سنة)	١,٢٠١	٠,٢١١	٥,٦٨٩	٠,٠٠٠

عدد المشاهدات = ٥٠، معامل التحديد R^2 = ٠,٩٦؛ معامل التحديد المعدل $\bar{R}^2$=٠,٩٦، الانحراف المعياري للتقدير s = ١,٠٩٨ ؛ درجات الحرية = ٤٧

٨-٢-٢ عرض نتائج نموذج الانحدار الخطي في شكل معادلة:

يعرض بعض الباحثين نتائج نموذج الانحدار في شكل معادلة. فمثلاً يتم عرض نتائج نموذج انحدار وزن الطفل (weight) على كل من الطول (height) والعمر (age) المستعرضة في الإطار رقم (٨-٣) كما يلي (Gujarati and Porter, 2009, p.129 :)

$$\begin{aligned}
\widehat{weight} &= -2.182 + 0.125\times height + 1.201\times age \\
se &= (0.976) \quad (0.018) \quad (0.211) \qquad R^2 = 0.962 \\
t &= (-2.235) \quad (6.745) \quad (5.689) \qquad df = 47 \\
p &= (0.030) \quad (0.000) \quad (0.000) \qquad F_{2,47} = 597.492
\end{aligned} \tag{8-2}$$

حيث إن se الخطأ المعياري وt قيمة توزيع t وp قيمة الاحتمال، وdf درجات الحرية، وF قيمة توزيع F

ووفقاً (Gupta, 2010) يتم عرض نتائج نموذج الانحدار الخطي في شكل معادلة كالتالي:

$$\begin{aligned}
\widehat{weight} &= -2.182 + 0.125\times height + 1.201\times age \\
t &= (-2.235)^{*} \quad (6.745)^{***} \quad (5.689)^{***} \\
R^2 &= 0.962 \\
\text{Adjusted } R^2 &= 0.961 \\
F(2,47) &= 597.49^{***} \\
n &= 50
\end{aligned} \tag{8-3}$$

*** مستوى دلالة/معنوية (٠,٠٠١)، ** مستوى دلالة/معنوية (٠,٠١)، * مستوى دلالة/معنوية (٠,٠٥)

٨-٢-٣ عرض نتائج نموذج الانحدار الخطي في شكل نص وفق نظام جمعية علم النفس الأمريكية:

بالإضافة إلى طريقة عرض نتائج نموذج الانحدار الخطي في شكل جدول أو معادلة، هناك طرق مختلفة لعرض نتائج نموذج الانحدار في شكل نصي ضمن نتائج الدراسة أو البحث. ويجب الإشارة هنا إلى أن العرض النصي جزء مكمل للعرض الجدولي أو عرض المعادلة. ووفقاً لجمعية علم النفس الأمريكية لا توجد طريقة محددة لعرض نتائج نموذج الانحدار في شكل نصي في متن البحث أو التقرير. وفيما يلي مثال لطريقة العرض النصي لنتائج نموذج انحدار وزن الطفل المستعرضة الإطارات (٨-٣، ٨-٥، ٨-٦) وفق نظام جمعية علم النفس الأمريكية (Lomax and Hahs-Vaughn, 2012).

"توضح نتائج نموذج الانحدار الخطي المتعدد أن نسبة كبيرة من التباين في وزن الطفل قد تم تفسيرها بواسطة متغيري الطول والعمر $\left(F_{(2,47)} = 597.492, p < .001\right)$. كما توضح النتائج أن معامل الطول (٠,١٢٥) يختلف عن الصفر بمستوى معنوي (t = 6.745, p < .001)، مما يشير إلى أن زيادة سم واحد في طول الطفل يسهم في زيادة الوزن بـ (١٢٥) جرام بافتراض ثبات العمر. كما أن معامل العمر البالغ (١,٢٠١) يختلف عن الصفر بمستوى معنوي (t = 5.689, p < .001) مما يوضح أن زيادة العمر بسنة واحدة تسهم في زيادة وزن الطفل بـ (١٢٠١) جرام بافتراض ثبات الطول. ويتضح من معامل التحديد (٠,٩٦٢) أن (٩٦,٢%) من التباين في وزن الطفل قد جرى تفسيره بواسطة طول وعمر الطفل".

الملاحق

ملحق (أ)
التوزيع الطبيعي والتوزيعات المتصلة الأخرى المرتبطة بها

يتناول هذا الملحق بإيجاز بعض التوزيعات المتصلة التي لها دور رئيسي في تطور النماذج الإحصائية الخطية وهي التوزيع الطبيعي، توزيع مربع كاي، توزيع 't' و توزيع 'F'.

أ- ١ التوزيع الطبيعي (The Normal Distribution):

التوزيع الطبيعي هو أكثر التوزيعات الاحتمالية استخداماً في التحليل الإحصائي لأنه يمثل كثيراً من الظواهر الطبيعية. أوزان، أعمار وأطوال الأشخاص، مستوي ضغط الدم، درجات التلميذ في أحد الاختبارات، الدخل، الاستهلاك، ... الخ أمثلة قليلة من متغيرات لا حصرٍ لها تتبع التوزيع الطبيعي. ويسمى التوزيع الطبيعي بتوزيع جاوس نسبة إلى مكتشفه كارل جاوس. ويعرف هذا التوزيع بواسطة دالة كثافة الاحتمال (Probability Density Function (PDF التي تأخذ الصيغة التالية:

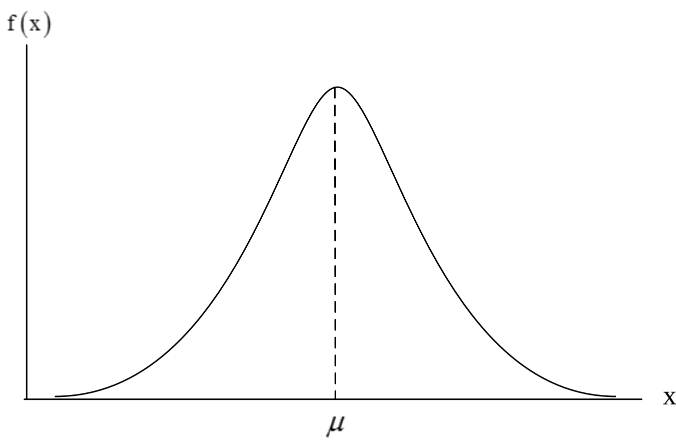
$$f\ (x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } -\infty < x < \infty \ , \ \sigma > 0$$

حيث أن X المتغير العشوائي المتصل الطبيعي ويأخذ قيماً ما بين سالب ما لانهاية إلى موجب ما لا نهاية ($-\infty < x < \infty$)، و μ و σ هما معلمتي التوزيع ويمثلان الوسط الحسابي وتباين التوزيع على التوالي و e قيمة ثابتة تساوي تقريباً (e=2.71828)، و π قيمة ثابتة تساوي تقريباً (3.1459)

بعض خصائص التوزيع الطبيعي:

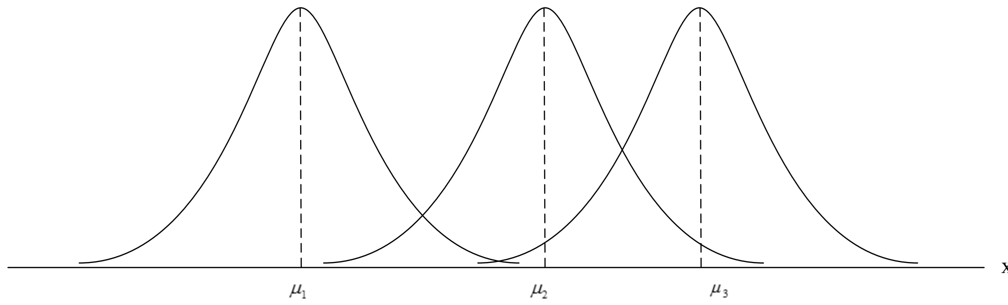
١) منحنى الدالة ذو قمة واحدة يشبه شكل الجرس ومتماثل حول الوسط الحسابي μ – أنظر الشكل (أ-١)، ومساحة المنطقة الواقعة بين هذا المنحنى والمحور X تساوي الواحد الصحيح، أي:

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1$$

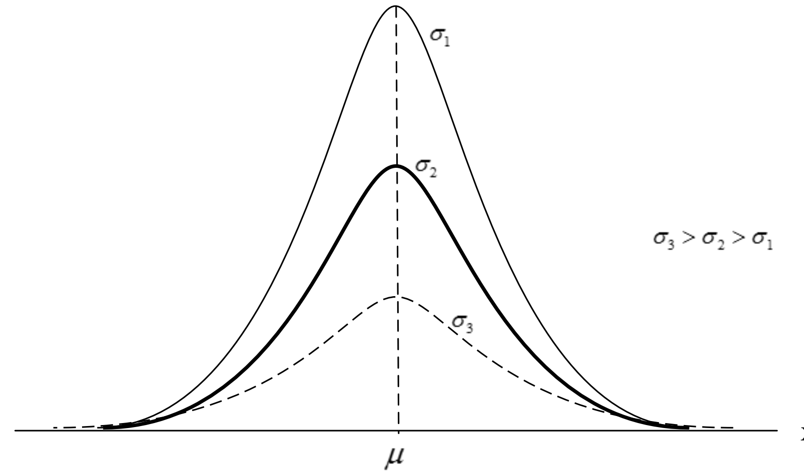


شكل (أ-١): منحنى التوزيع الطبيعي

٢) المعلمة μ (الوسط الحسابي) تحدد الموضع بينما المعلمة σ^2 (التباين) تحدد شكل التوزيـع (انظـر الشـكلين أ-٢ وأ-٣).



شكل (أ-٢): ثلاثة منحنيات لتوزيع طبيعي لثلاثة قيم مختلفة للوسط الحسابي وتباين ثابت



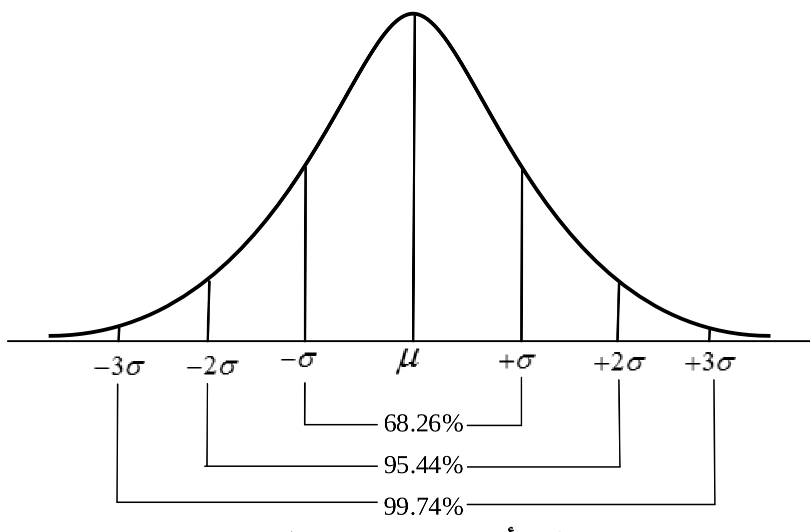
شكل (أ-٣): ثلاثة منحنيات لتوزيع طبيعي لثلاثة قيم مختلفة للتباين ووسط حسابي ثابت

٣) توزيع المساحة المحصورة تحت المنحني على المحور الأفقي على النحو التالي:

٦٨,٣% من قيم المتغير العشوائي المتصل تقع ما بين $\mu+\sigma$ و $\mu-\sigma$

٩٥,٤% من قيم المتغير العشوائي المتصل تقع ما بين $\mu+2\sigma$ و $\mu-2\sigma$

٩٩,٧% من قيم المتغير العشوائي المتصل تقع ما بين $\mu+3\sigma$ و $\mu-3\sigma$



شكل (أ-٤): توزيع المنحنى الطبيعي

٤) التوزيع الطبيعي المعياري **(The Standard Normal Distribution):**

إذا كان X متغير عشوائي طبيعي بوسط حسابي يساوي μ وتباين يساوي σ^2، فإن المتغير **Z** حيث

$$z = \frac{x-\mu}{\sigma} \sim N(0,1)$$

يتبع أيضاً التوزيع الطبيعي بوسط حسابي يساوى الصفر وانحراف معياري أو تباين يساوي الواحد الصحيح. ويعرف هذا المتغير بالمتغير الطبيعي المعياري. وللتوزيع الطبيعي المعياري نفس خصائص التوزيع الطبيعي باستثناء أن الوسط الحسابي يساوي صفر والتباين يساوى واحد صحيح. ويلاحظ أن الغالبية العظمى من قيم المتغير الطبيعي المعياري تقع بين −٣ و+٣، أي أن أعلى قيمة يأخذها المتغير هي ٣ تقريباً وأقل قيمة هي −٣ تقريباً (الشكل أ-٤). وباستبدال Z في دالة كثافة التوزيع الطبيعي نحصل على دالة كثافة المتغير الطبيعي المعياري كما يلي:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

وفيما يلي بعض القيم المعيارية والاحتمالات المتجمعة المناظرة الواسعة الاستخدام:

قيم Z المعيارية	الاحتمال المتجمع (من $-\infty$ إلى Z)
١,٢٨٢	٠,٩٠٠
١,٦٤٥	٠,٩٥٠
١,٩٦	٠,٩٧٥
٢,٣٢٦	٠,٩٩٥
٢,٥٧٥	٠,٩٩٥

٥) من خصائص التوزيع الطبيعي، إنه إذا ما وجدت علاقة خطية بين متغير يتبع التوزيع الطبيعي وبين متغير آخـر، فإن المتغير الآخر يتبع التوزيع الطبيعي أيضاً.

مثال:

إذا كانت $Y_1, Y_2, \ldots, Y_m$ متغيرات طبيعية مستقلة عنها البعض ولديها التوزيع التالي:

$$Y_i \sim N(\mu_i, \sigma_i^2) \qquad \text{for } i = 1,2,\ldots,m$$

وإن W متغير عشوائي عبارة عن تركيب خطي للمتغيرات $Y_1, Y_2, \ldots, Y_m$ ، أي:

$$W = a_1Y_1 + a_2Y_2 + \ldots + a_mY_m$$

حيث أن a_i ثابت، فإن المتغير W يتبع التوزيع الطبيعي أيضاً بوسط حسابي $E(W) = a_1\mu_1 + a_2\mu_2 + \ldots + a_m\mu_m$

وتباين $\text{var}(W) = a_1^2\sigma^2{}_1 + a_2^2\sigma^2{}_2 + \ldots + a_m^2\sigma_m^2$ ؛ أو اختصارا

$$W = \sum_{i=1}^{m} a_iY_i \sim N\left(\sum_{i=1}^{m} a_i\mu_i, \sum_{i=1}^{m} a_i^2\sigma_i^2\right)$$

أ-٢ توزيع مربع كاي (Chi-squared Distribution)

يقال لمتغير عشوائي متصل (X) أنه يتبع توزيع مربع كاي (χ^2) بدرجات حرية قدرها (m)،إذا كانت دالـة كثافـة احتمال المتغير (X) تأخذ الصيغة التالية:

$$f(x) = \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} x^{\frac{(m-2)}{2}} e^{\frac{-x}{2}} \qquad \text{for } x > 0$$

حيث إن:

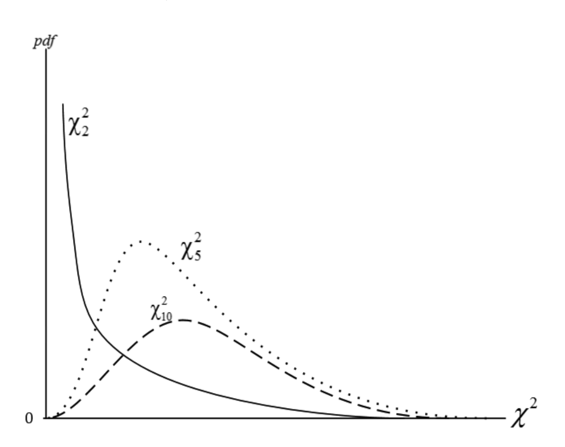
$\Gamma(\frac{m}{2})=(\frac{m}{2}-1)!$ إذا كان m عدد زوجي و $\Gamma\left(\frac{m}{2}\right)=\left(\frac{m}{2}-1\right)\left(\frac{m}{2}-2\right)\left(\frac{m}{2}-3\right)\ldots\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}$ إذا كان فردياً.

خواص توزيع مربع كاي:

١) يعتمد توزيع مربع كاي اعتماداً كاملاً على درجات الحرية (m) المعلمة الوحيدة للتوزيع.

٢) توزيع مربع كاي توزيع ملتو نحو اليمين (Right-skew) خاصة إذا كان عدد درجات الحرية قليلاً.

٣) يقترب توزيع مربع كاي إلى التماثل – أي يقترب من التوزيع الطبيعي- كلما زادت درجات الحرية (انظر الشكل (أ-٥)).



شكل (أ-٥): دالة كثافة توزيع مربع كاي عند درجات حرية ٢، ٥، ١٠.

٤) إذا كان المتغير X يتبع توزيع مربع كاي بدرجات حرية (m) ، فإن الوسط الحسابي للمتغير هو:

$$E(x)=m$$

وتباين المتغير X هو:

$$\text{var}(x)=2m$$

٥) إذا كانت $Z_1,Z_2,\ldots,Z_m$ متغيرات طبيعية معيارية مستقلة (Independent Standardized Variables)، أي:

$$Z_1,\ Z_2,\ \ldots,\ Z_m \sim N\ (0\ ,\ 1)$$

فإن المتغير العشوائي Z حيث

$$Z=\sum_{i=1}^{m} Z_i^2 \quad \sim \quad \chi_m^2$$

يتبع توزيع مربع كاي بدرجات حرية (m).

٦) إذا كانت $x_1, x_2,\ldots,x_n$ عينة عشوائية من توزيع طبيعي وسطه الحسابي μ وتباينه σ^2 فإن Q حيث

$$Q = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2} = \frac{(n-1)\ s^2}{\sigma^2} \quad \sim \chi^2_{(n-1)}$$

يتبع توزيع مربع كاي بدرجات حرية (n-1). حيث إن n حجم العينة و s^2 تباين العينة و σ^2 تباين المجتمع.

٧) **نظرية النهاية المركزية (Central limit theorem):** إذا كانت $X_1, X_2,\ldots, X_n$ متغيرات عشوائية مستقلة وتوزيعها مماثل بوسط حسابي $E(X_i)=\mu$ وتباين $\operatorname{var}(X_i)=\sigma^2<\infty$ ، فإن القيمة

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

تؤول إلى التوزيع الطيعي المعياري كلما كان حجم العينة كبيراً $(n\rightarrow\infty)$.

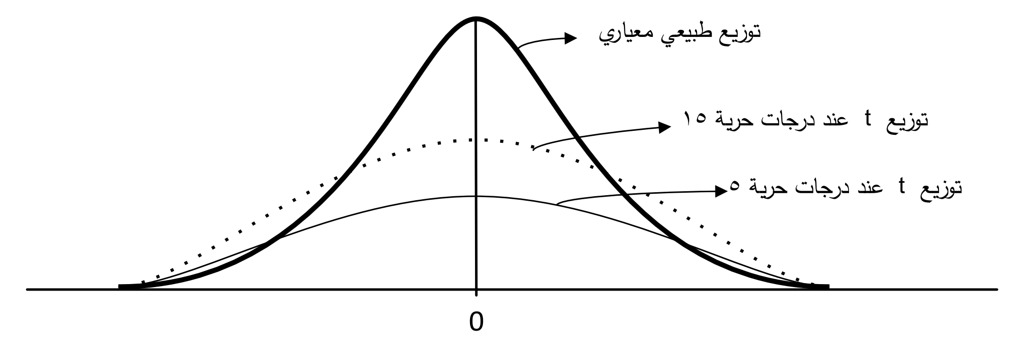
أ-٣ توزيع 't' (t Distribution):

يقال لمتغير عشوائي متصل X أن له توزيع 't' إذا كانت دالة كثافة احتماله تأخذ الصيغة التالية:

$$f\,(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\ \Gamma(\frac{n}{2})} \cdot \frac{1}{(1+\frac{x^2}{n})^{(n+1)/2}} \qquad \text{for} \quad -\infty < x < +\infty$$

خصائص توزيع 't':

- توزيع متصل متماثل حول الصفر ولكنه منبسط عن التوزيع الطبيعي، ولهذا فإن جزءاً أكبر من مساحته تقع عن الأطراف (انظر الشكل رقم (أ-٦)).



شكل (أ-٦): توزيع 't'

- له معلمة واحدة وهي n -درجات الحرية- تحدد شكل منحنى التوزيع. ويقترب توزيع 't' من التوزيع الطبيعي المعياري كلما زادت درجات الحرية واقترب من لانهاية.

- الوسط الحسابي للتوزيع يساوي صفر وتباينه يساوي $\sigma^2 = \left(\frac{n}{n-2}\right)$ ، حيث n أكبر من ٢.

- افترض أن W_1 وW_2 متغيران مستقلان يتبعان التوزيع الطبيعي المعياري وتوزيع مربع كاي بدرجات حرية n على التوالي($W_1 \sim N(0,1)$ ؛ $W_2 \sim \chi^2_n$)، فإن المتغير العشوائي T حيث

$$T = \frac{W_1}{\sqrt{\frac{W_2}{n}}} \sim t_n$$

يتبع توزيع t بدرجات حرية تساوي n.

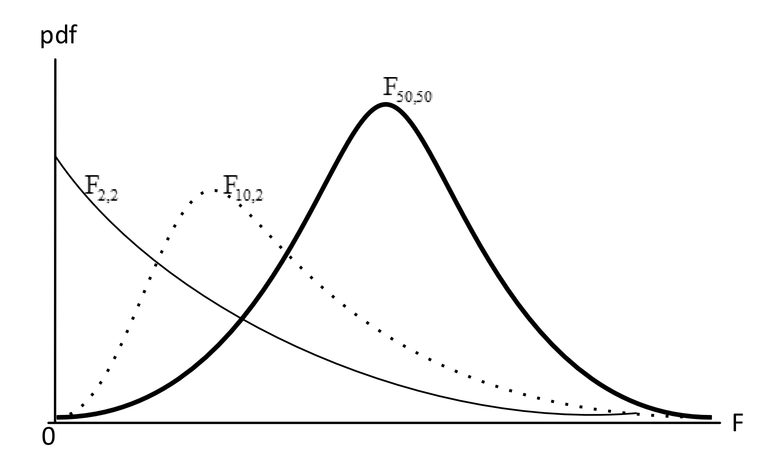
أ-٤ توزيع F:

يقال لمتغير عشوائي متصل X له توزيع F إذا كانت دالة كثافة احتماله تأخذ الصيغة التالية:

$$f(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{(n_1-2)}{2}}\left(1+\frac{n_1}{n_2}\right)^{-\left(\frac{n_1+n_2}{2}\right)} \qquad 0 < x < +\infty$$

خصائص توزيع F:

- توزيع متصل ملتو نحو اليمين.
- له معلمتان n_1 و n_2 تحدد منحني التوزيع وتسمى بدرجات الحرية.
- يقترب التوزيع إلى التوزيع كلما زادت درجات الحرية n_1 و n_2 (الشكل أ-٧).



شكل (أ-٧): منحنيان توزيع 'F'لقيم مختلفة لدرجات الحرية

- الوسط الحسابي للتوزيع يساوى $E(x) = \frac{n_2}{(n_2-2)}$ وتباينه يساوي $var(x) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)(n_2-4)}$

- افترض أن Z_1 و Z_2 متغيران عشوائيان مستقلان يتبعان توزيع مربع كاي بـدرجات حريـة n_1 وn_2 عـلى التـوالي فإن المتغير F حيث

$$F = \frac{Z_1/n_1}{Z_2/n_2} \sim F_{n_1,n_2}$$

له توزيع 'F' بدرجات حرية n_1 و n_2.

- مربع المتغير العشوائي الذي يتبع توزيع 't' بدرجات حرية n له توزيع F بدرجات حرية ١ و n، أي:

$$t_n^2 = F_{1,n}$$

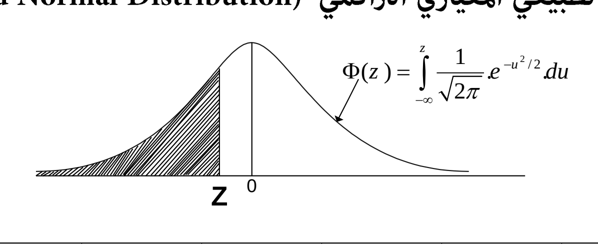
- إذا كان s_1^2 و s_2^2 تباينان لعينتين عشوائيتين مستقلتين حجمهما n_1 و n_2 على التوالي سحبتا من توزيعين طبيعيـين مختلفين تباينهما σ_1^2 و σ_2^2 على التوالي، فإن المتغير F حيث

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

يتبع توزيع F بدرجتي حرية (n_1-1) و (n_2-1) ويستخدم التوزيع لاختبار تساوي تباينين مجتمعين.

ملحق "ب"
جداول إحصائية

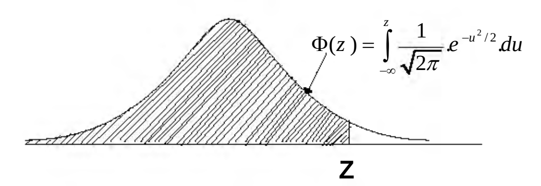
ملحق ١: جدول التوزيع الطبيعي المعياري التراكمي* (Cumulative Standard Normal Distribution)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0-	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

*** تم حساب هذه القيم باستخدام برنامج إكسل (Excel 2010)**

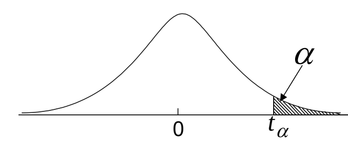
تابع التوزيع الطبيعي المعياري التراكمي*



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

*** تم حساب هذه القيم باستخدام برنامج إكسل (Excel 2010)**

القيم الحرجة لتوزيع t



درجات الحرية	مستوى المعنوية: جانب واحد	0.100	0.050	0.025	0.010	0.005
	ذو جانبين	0.200	0.100	0.050	0.020	0.010
	1	3.078	6.314	12.706	31.821	63.656
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787

درجات الحرية	مستوى المعنوية	جانب واحد	0.100	0.050	0.025	0.010	0.005
		ذو جانبين	0.200	0.100	0.050	0.020	0.010
درجات الحرية		26	1.315	1.706	2.056	2.479	2.779
		27	1.314	1.703	2.052	2.473	2.771
		28	1.313	1.701	2.048	2.467	2.763
		29	1.311	1.699	2.045	2.462	2.756
		40	1.303	1.684	2.021	2.423	2.704
		50	1.299	1.676	2.009	2.403	2.678
		60	1.296	1.671	2.000	2.390	2.660
		70	1.294	1.667	1.994	2.381	2.648
		80	1.292	1.664	1.990	2.374	2.639
		90	1.291	1.662	1.987	2.368	2.632
		100	1.290	1.660	1.984	2.364	2.626
		120	1.289	1.658	1.980	2.358	2.617
		∞	1.282	1.645	1.960	2.326	2.576

تم حساب هذه القيم باستخدام برنامج إكسل (Excel 2010)

القيم الحرجة لتوزيع F

(مستوى المعنوية ٥%)

درجات الحرية للمقام	درجات الحرية للبسط														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161.	199.	215.	224.	230.	234.	236.	238.	240.	241.	243.	243.	244.	245.	246.
2	18.5	19	19.1	19.2	19.3	19.3	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.7
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6	5.96	5.94	5.91	5.89	5.87	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.7	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06	4.03	4	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.6	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.5	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.1	3.07	3.05	3.03	3.01
10	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85
11	4.84	3.98	3.59	3.36	3.2	3.09	3.01	2.95	2.9	2.85	2.82	2.79	2.76	2.74	2.72
12	4.75	3.89	3.49	3.26	3.11	3	2.91	2.85	2.8	2.75	2.72	2.69	2.66	2.64	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.6	2.58	2.55	2.53
14	4.6	3.74	3.34	3.11	2.96	2.85	2.76	2.7	2.65	2.6	2.57	2.53	2.51	2.48	2.46
15	4.54	3.68	3.29	3.06	2.9	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.4
20	4.35	3.49	3.1	2.87	2.71	2.6	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.2
25	4.24	3.39	2.99	2.76	2.6	2.49	2.4	2.34	2.28	2.24	2.2	2.16	2.14	2.11	2.09
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2	1.97	1.95	1.92
50	4.03	3.18	2.79	2.56	2.4	2.29	2.2	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87
60	4	3.15	2.76	2.53	2.37	2.25	2.17	2.1	2.04	1.99	1.95	1.92	1.89	1.86	1.84
70	3.98	3.13	2.74	2.5	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84	1.81
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2	1.95	1.91	1.88	1.84	1.82	1.79
90	3.95	3.1	2.71	2.47	2.32	2.2	2.11	2.04	1.99	1.94	1.9	1.86	1.83	1.8	1.78
100	3.94	3.09	2.7	2.46	2.31	2.19	2.1	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77

تابع القيم الحرجة لتوزيع F
(مستوى المعنوية ٥%)

درجات الحرية للمقام \ درجات الحرية للبسط	20	25	30	40	50	60	70	80	90	100	120	140	160	180	200
1	248.0	249.3	250.1	251.1	251.8	252.2	252.5	252.7	252.9	253.0	253.3	253.4	253.5	253.6	253.7
2	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	8.66	8.63	8.62	8.59	8.58	8.57	8.57	8.56	8.56	8.55	8.55	8.55	8.54	8.54	8.54
4	5.80	5.77	5.75	5.72	5.70	5.69	5.68	5.67	5.67	5.66	5.66	5.65	5.65	5.65	5.65
5	4.56	4.52	4.50	4.46	4.44	4.43	4.42	4.41	4.41	4.41	4.40	4.39	4.39	4.39	4.39
6	3.87	3.83	3.81	3.77	3.75	3.74	3.73	3.72	3.72	3.71	3.70	3.70	3.70	3.69	3.69
7	3.44	3.40	3.38	3.34	3.32	3.30	3.29	3.29	3.28	3.27	3.27	3.26	3.26	3.25	3.25
8	3.15	3.11	3.08	3.04	3.02	3.01	2.99	2.99	2.98	2.97	2.97	2.96	2.96	2.95	2.95
9	2.94	2.89	2.86	2.83	2.80	2.79	2.78	2.77	2.76	2.76	2.75	2.74	2.74	2.73	2.73
10	2.77	2.73	2.70	2.66	2.64	2.62	2.61	2.60	2.59	2.59	2.58	2.57	2.57	2.57	2.56
11	2.65	2.60	2.57	2.53	2.51	2.49	2.48	2.47	2.46	2.46	2.45	2.44	2.44	2.43	2.43
12	2.54	2.50	2.47	2.43	2.40	2.38	2.37	2.36	2.36	2.35	2.34	2.33	2.33	2.33	2.32
13	2.46	2.41	2.38	2.34	2.31	2.30	2.28	2.27	2.27	2.26	2.25	2.25	2.24	2.24	2.23
14	2.39	2.34	2.31	2.27	2.24	2.22	2.21	2.20	2.19	2.19	2.18	2.17	2.17	2.16	2.16
15	2.33	2.28	2.25	2.20	2.18	2.16	2.15	2.14	2.13	2.12	2.11	2.11	2.10	2.10	2.10
20	2.12	2.07	2.04	1.99	1.97	1.95	1.93	1.92	1.91	1.91	1.90	1.89	1.88	1.88	1.88
25	2.01	1.96	1.92	1.87	1.84	1.82	1.81	1.80	1.79	1.78	1.77	1.76	1.75	1.75	1.75
30	1.93	1.88	1.84	1.79	1.76	1.74	1.72	1.71	1.70	1.70	1.68	1.68	1.67	1.66	1.66
40	1.84	1.78	1.74	1.69	1.66	1.64	1.62	1.61	1.60	1.59	1.58	1.57	1.56	1.55	1.55
50	1.78	1.73	1.69	1.63	1.60	1.58	1.56	1.54	1.53	1.52	1.51	1.50	1.49	1.49	1.48
60	1.75	1.69	1.65	1.59	1.56	1.53	1.52	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.44
70	1.72	1.66	1.62	1.57	1.53	1.50	1.49	1.47	1.46	1.45	1.44	1.42	1.42	1.41	1.40
80	1.70	1.64	1.60	1.54	1.51	1.48	1.46	1.45	1.44	1.43	1.41	1.40	1.39	1.38	1.38
90	1.69	1.63	1.59	1.53	1.49	1.46	1.44	1.43	1.42	1.41	1.39	1.38	1.37	1.36	1.36
100	1.68	1.62	1.57	1.52	1.48	1.45	1.43	1.41	1.40	1.39	1.38	1.36	1.35	1.35	1.34

تابع القيم الحرجة لتوزيع F

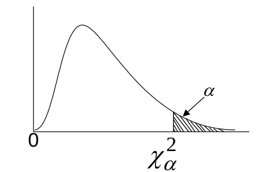
(مستوى المعنوية ١%)

درجات الحرية للمقام	درجات الحرية للبسط														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.1	6022.5	6055.9	6083.3	6106.3	6125.9	6142.7	6157.3
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.42	99.43	99.43
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05	26.98	26.92	26.87
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.31	14.25	14.20
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.82	9.77	9.72
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56	2.51	2.46	2.42
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35	2.31
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29	2.24
10	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.27	2.22

تابع القيم الحرجة لتوزيع F
(مستوى المعنوية ١%)

درجات الحرية للمقام	درجات الحرية للبسط														
	20	25	30	40	50	60	70	80	90	100	120	140	160	180	200
1	6208.7	6239.8	6260.7	6286.8	6302.5	6313.0	6320.6	6326.2	6330.6	6334.1	6339.4	6343.2	6346.0	6348.2	6350.0
2	99.45	99.46	99.47	99.47	99.48	99.48	99.48	99.49	99.49	99.49	99.49	99.49	99.49	99.49	99.49
3	26.69	26.58	26.50	26.41	26.35	26.32	26.29	26.27	26.25	26.24	26.22	26.21	26.20	26.19	26.18
4	14.02	13.91	13.84	13.75	13.69	13.65	13.63	13.61	13.59	13.58	13.56	13.54	13.53	13.53	13.52
5	9.55	9.45	9.38	9.29	9.24	9.20	9.18	9.16	9.14	9.13	9.11	9.10	9.09	9.08	9.08
6	7.40	7.30	7.23	7.14	7.09	7.06	7.03	7.01	7.00	6.99	6.97	6.96	6.95	6.94	6.93
7	6.16	6.06	5.99	5.91	5.86	5.82	5.80	5.78	5.77	5.75	5.74	5.72	5.72	5.71	5.70
8	5.36	5.26	5.20	5.12	5.07	5.03	5.01	4.99	4.97	4.96	4.95	4.93	4.92	4.92	4.91
9	4.81	4.71	4.65	4.57	4.52	4.48	4.46	4.44	4.43	4.41	4.40	4.39	4.38	4.37	4.36
10	4.41	4.31	4.25	4.17	4.12	4.08	4.06	4.04	4.03	4.01	4.00	3.98	3.97	3.97	3.96
11	4.10	4.01	3.94	3.86	3.81	3.78	3.75	3.73	3.72	3.71	3.69	3.68	3.67	3.66	3.66
12	3.86	3.76	3.70	3.62	3.57	3.54	3.51	3.49	3.48	3.47	3.45	3.44	3.43	3.42	3.41
13	3.66	3.57	3.51	3.43	3.38	3.34	3.32	3.30	3.28	3.27	3.25	3.24	3.23	3.23	3.22
14	3.51	3.41	3.35	3.27	3.22	3.18	3.16	3.14	3.12	3.11	3.09	3.08	3.07	3.06	3.06
15	3.37	3.28	3.21	3.13	3.08	3.05	3.02	3.00	2.99	2.98	2.96	2.95	2.94	2.93	2.92
20	2.94	2.84	2.78	2.69	2.64	2.61	2.58	2.56	2.55	2.54	2.52	2.50	2.49	2.49	2.48
25	2.70	2.60	2.54	2.45	2.40	2.36	2.34	2.32	2.30	2.29	2.27	2.26	2.25	2.24	2.23
30	2.55	2.45	2.39	2.30	2.25	2.21	2.18	2.16	2.14	2.13	2.11	2.10	2.09	2.08	2.07
40	2.37	2.27	2.20	2.11	2.06	2.02	1.99	1.97	1.95	1.94	1.92	1.90	1.89	1.88	1.87
50	2.27	2.17	2.10	2.01	1.95	1.91	1.88	1.86	1.84	1.82	1.80	1.79	1.77	1.76	1.76
60	2.20	2.10	2.03	1.94	1.88	1.84	1.81	1.78	1.76	1.75	1.73	1.71	1.70	1.69	1.68
70	2.15	2.05	1.98	1.89	1.83	1.78	1.75	1.73	1.71	1.70	1.67	1.65	1.64	1.63	1.62
80	2.12	2.01	1.94	1.85	1.79	1.75	1.71	1.69	1.67	1.65	1.63	1.61	1.60	1.59	1.58
90	2.09	1.99	1.92	1.82	1.76	1.72	1.68	1.66	1.64	1.62	1.60	1.58	1.57	1.55	1.55
100	2.07	1.97	1.89	1.80	1.74	1.69	1.66	1.63	1.61	1.60	1.57	1.55	1.54	1.53	1.52

ملحق ٣: جدول توزيع χ^2



درجات الحرية	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
1	7.879	6.635	5.024	3.841	2.706	0.016	0.004	0.001	0.000	0.000
2	10.597	9.210	7.378	5.991	4.605	0.211	0.103	0.051	0.020	0.010
3	12.838	11.345	9.348	7.815	6.251	0.584	0.352	0.216	0.115	0.072
4	14.860	13.277	11.143	9.488	7.779	1.064	0.711	0.484	0.297	0.207
5	16.750	15.086	12.832	11.070	9.236	1.610	1.145	0.831	0.554	0.412
6	18.548	16.812	14.449	12.592	10.645	2.204	1.635	1.237	0.872	0.676
7	20.278	18.475	16.013	14.067	12.017	2.833	2.167	1.690	1.239	0.989
8	21.955	20.090	17.535	15.507	13.362	3.490	2.733	2.180	1.647	1.344
9	23.589	21.666	19.023	16.919	14.684	4.168	3.325	2.700	2.088	1.735
10	25.188	23.209	20.483	18.307	15.987	4.865	3.940	3.247	2.558	2.156
11	26.757	24.725	21.920	19.675	17.275	5.578	4.575	3.816	3.053	2.603
12	28.300	26.217	23.337	21.026	18.549	6.304	5.226	4.404	3.571	3.074
13	29.819	27.688	24.736	22.362	19.812	7.041	5.892	5.009	4.107	3.565
14	31.319	29.141	26.119	23.685	21.064	7.790	6.571	5.629	4.660	4.075
15	32.801	30.578	27.488	24.996	22.307	8.547	7.261	6.262	5.229	4.601
16	34.267	32.000	28.845	26.296	23.542	9.312	7.962	6.908	5.812	5.142
17	35.718	33.409	30.191	27.587	24.769	10.085	8.672	7.564	6.408	5.697
18	37.156	34.805	31.526	28.869	25.989	10.865	9.390	8.231	7.015	6.265
19	38.582	36.191	32.852	30.144	27.204	11.651	10.117	8.907	7.633	6.844
20	39.997	37.566	34.170	31.410	28.412	12.443	10.851	9.591	8.260	7.434
21	41.401	38.932	35.479	32.671	29.615	13.240	11.591	10.283	8.897	8.034
22	42.796	40.289	36.781	33.924	30.813	14.041	12.338	10.982	9.542	8.643
23	44.181	41.638	38.076	35.172	32.007	14.848	13.091	11.689	10.196	9.260
24	45.558	42.980	39.364	36.415	33.196	15.659	13.848	12.401	10.856	9.886
25	46.928	44.314	40.646	37.652	34.382	16.473	14.611	13.120	11.524	10.520
26	48.290	45.642	41.923	38.885	35.563	17.292	15.379	13.844	12.198	11.160
27	49.645	46.963	43.195	40.113	36.741	18.114	16.151	14.573	12.878	11.808
28	50.994	48.278	44.461	41.337	37.916	18.939	16.928	15.308	13.565	12.461
29	52.335	49.588	45.722	42.557	39.087	19.768	17.708	16.047	14.256	13.121
30	53.672	50.892	46.979	43.773	40.256	20.599	18.493	16.791	14.953	13.787
40	66.766	63.691	59.342	55.758	51.805	29.051	26.509	24.433	22.164	20.707
50	79.490	76.154	71.420	67.505	63.167	37.689	34.764	32.357	29.707	27.991
60	91.952	88.379	83.298	79.082	74.397	46.459	43.188	40.482	37.485	35.534
70	104.215	100.425	95.023	90.531	85.527	55.329	51.739	48.758	45.442	43.275
80	116.321	112.329	106.629	101.879	96.578	64.278	60.391	57.153	53.540	51.172
90	128.299	124.116	118.136	113.145	107.565	73.291	69.126	65.647	61.754	59.196
100	140.170	135.807	129.561	124.342	118.498	82.358	77.929	74.222	70.065	67.328

تم حساب هذه القيم باستخدام برنامج إكسل (Excel 2010)

جدول إحصاء ديربن-واتسون

(مستوى المعنوية0.05)

عدد المشاهدات (n)	p*=1		p=2		p=3		p=4		p=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	1.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79

عدد المشاهدات (n)	$p^*=1$		p=2		p=3		p=4		p=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

* p = عدد المتغيرات المستقلة

المصدر:

Durbin, J. and Watson, G. S, "Testing for Serial Correlation in Least Squares Regression. II," *Biometrika* 38 (1951), pp. 159-178

جدول إحصاء ديربن-واتسون

(مستوى المعنوية 0.01)

عدد المشاهدات (n)	$p^*=1$		p=2		p=3		p=4		p=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58

عدد المشاهدات (n)	p*=1		p=2		p=3		p=4		p=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

* p = عدد المتغيرات المستقلة

القيم الحرجة لاختبار تبعية المتغيرات للتوزيع الطبيعي

(Critical Values for the Normal Scores Correlation Test for Normality)

حجم العينة	مستوى المعنوية (Significance Level)		
	0.01	**0.05**	**0.10**
10	0.879	0.917	0.935
15	0.91	0.938	0.951
20	0.928	0.951	0.96
25	0.94	0.958	0.967
30	0.949	0.964	0.971
40	0.958	0.972	0.977
50	0.966	0.976	0.981
60	0.971	0.98	0.983
70	0.975	0.982	0.985
80	0.978	0.984	0.987
90	0.98	0.986	0.988
100	0.982	0.987	0.989
200	0.99	0.993	0.994
300	0.993	0.995	0.996
400	0.995	0.996	0.997
500	0.996	0.997	0.998
1000	0.998	0.998	0.999

المصدر : Weiss, 2002 p. A24

ملحق "ج"
قائمة بالمصطلحات المستخدمة

قائمة بالمصطلحات المستخدمة
Glossary

A

Acceptance region	منطقة القبول
Adjusted R-square	معامل التحديد المعدل
Alternative hypothesis	الفرض البديل
Analysis of variance table	جدول تحليل التباين
Autocorrelation	ارتباط ذاتي
Autocovariance	التغاير الذاتي
Autoregressive model	نموذج انحدار ذاتي

B

Backward selection method	طريقة الحذف إلى الخلف
Bias	تحيز
Biasing constant	ثابت التحيز
Binary variable	متغير ثنائي
Binomial distribution	توزيع ذي الحدين
BLUE (best linear unbiased estimators)	أفضل مقدرات خطية غير متحيزة
Box-Cox transformation	تحويلة بوكس-كوكس
Boxplot	الرسم الصندوقي
Breush-Pagan test	اختبار بروش-باقان

C

Central limit theorem	نظرية النهاية المركزية
Chi-square distribution	توزيع مربع كاي
Cochrane-Orcutt method	طريقة كوكرين-أوركت
Coding	ترميز
Coefficient	معامل
Coefficient of determination	معامل التحديد
Coefficient of partial determination	معامل التحديد الجزئي
Conditional probability function	دالة الاحتمال الشرطي
Confidence interval	فترة ثقة
Confidence limits	حدود ثقة

Consistency	اتساق
Consistent	متسق
Contrasts	مقارنات
Continuous distribution	توزيع مستمر/متصل
Continuous variable	متغير مستمر/متصل
Cook's distance measure	مقياس مسافة كوك
Correlation coefficient	معامل ارتباط
Correlation matrix	مصفوفة ارتباط
Correlation transformation	تحويلة الارتباط
Covariance	تغاير
Critical region	منطقة حرجة
Cross-sectional data	بيانات مقطعية
Cubic	تكعيبي
Cumulative normal function	دالة التوزيع الطبيعي التراكمي
Cumulative probability distribution	التوزيع الاحتمالي المتجمع

D

Data matrix	مصفوفة البيانات
Data transformation	تحويل البيانات
Degrees of freedom	درجات حرية
Deleted residuals	بواقي محذوفة
Density function	دالة كثافة
Dependent variable	متغير تابع
Determinant	محددة
Diagonal matrix	مصفوفة قطرية
Discrete data	بيانات متقطعة
Discrete distribution	توزيع متقطع
Dummy variable	متغير صوري
Dummy variables trap	مصيدة المتغيرات الصورية
Durbin-Watson statistic	إحصاء ديربن-واتسون

E

Efficiency	فعالية
Eigenvalues	قيم مميزة

Eigenvectors	المتجهات المميزة
Error Term	حد خطأ
Estimation	تقدير
Estimator	مقدر
Excel	برنامج اكسل
Expectation	توقع
Expected value	قيمة متوقعة
Explained variation	التباين المفسر
Explanatory variable	متغير تفسيري
Exponential model	نموذج أسي
Extra sum of squares	مجموع مربعات إضافي
Extreme value	قيمة متطرفة

F

F-distribution	توزيع F
F-statistic	إحصاء F
Factor analysis	التحليل العاملي
First-Order autocorrelation	ارتباط ذاتي من الرتبة الأولى
First-order autoregressive model	نموذج انحدار ذاتي من الرتبة الأولى
First differences	فروقات أولية
Forecasting	تنبؤ
Forward selection method	طريقة الاختيار إلى الأمام
Full model	نموذج كامل/تام

G

Gauss-Markov theorem	نظرية جاوس-ماركوف
General linear model	نموذج خطي عام
Generalized Differences Method	طريقة الفروقات المعممة
Generalized least squares	المربعات الصغرى المعممة
Generalized variance inflation factor	عامل تضخم التباين المعمم
Glejser's test	اختبار جليجسر
Goldfeld-Qundandt test	اختبار جولدفيلد-كواندت
Goodness of fit	جودة التوفيق

H

Hat matrix	مصفوفة القبعة
Heteroscedasticity	اختلاف التباين
Hinge spread	المدى الربيعي
Histogram	المدرج التكراري
Homoscedasticity	ثبات التباين
Hypothesis testing	اختبار الفرضية

I

Idempotent matrix	مصفوفة جامدة
Identity matrix	مصفوفة وحدة
Imperfect multicollinearity	ارتباط خطي غير تام
Independent variable	متغير مستقل
Inferential statistic	إحصاء استدلالي
Interaction	تفاعل
Interquartile range	المدى الربيعي
Interval scale	مقياس فتري
Inverse of the matrix	معكوس المصفوفة

J

Joint distribution	توزيع مشترك

L

Lack of fit test	**اختبار نقص المطابقة**
Lag	متباطئة
Lagged variable	متغير متباطئ
Least squares method	طريقة المربعات الصغرى
Level of significance	مستوى معنوية/دلالة
Linear combination	تركيب خطي
Linear regression model	نموذج انحدار خطي
Linear simultaneous systems of equations	نماذج المعادلات الخطية الآنية
Linearity in parameters	خطية في المعالم
Linearity in variables	خطية في المتغيرات
Logistic regression model	نموذج انحدار لوجستي

M

Mallows's statistic	إحصاء ملاوس
Matrix algebra	جبر المصفوفات
Maximum likelihood method	طريقة الترجيح الأعظم
Mean squared error	متوسط مربعات الخطأ
Minitab	برنامج مينيتاب للإحصاء
Missing value	قيمة مفقودة
Model Misspecification	خطأ توصيف النموذج
Multicollinearity	ارتباط خطي
Multiple comparison	مقارنات متعددة
Multiple correlation coefficient	معامل الارتباط المتعدد
Multiple linear regression model	نموذج انحدار خطي متعدد

N

Nominal scale	مقياس اسمي
Non-linear simultaneous systems of equations	نماذج المعادلات غير الخطية الآنية
Nonsingular matrix	مصفوفة غير مفردة
Normal distribution	توزيع طبيعي
Normal probability plot	رسم الاحتمال الطبيعي
Normal scores	الدرجات المعيارية
Null hypothesis	فرض عدم/فرض صفري
Null matrix	مصفوفة صفرية

O

Omission of relevant variables	عدم إدخال المتغيرات المستقلة ذات العلاقة
One-way analysis of variance	تحليل التباين في اتجاه واحد
Order condition	شرط الرتبة
Ordinal data	بيانات ترتيبية
Ordinary least squares (OLS)	المربعات الصغرى الاعتيادية
Orthogonal polynomials	الحدود المتعددة المتعامدة
Outlying observations	مشاهدات شاذة (متطرفة)

P

P-value	قيمة الاحتمال
Parameter	معلمة

Parsimony principle	مبدأ التبسيط / مبدأ الاختصار
Park's test	اختبار بارك
Partial Correlation Coefficient	معامل الارتباط الجزئي
Partial regression plot	رسم الانحدار الجزئي
Perfect multicollinearity	ارتباط خطي تام
Point estimation	تقدير النقطة
Poisson distribution	توزيع بواسون
Polynomial distributed lags	المتباطئات الموزعة ذات الحدود المتعددة
Polynomial regression	انحدار متعدد الحدود
Population	مجتمع
Power transformation	تحويلات القوة
Prediction Interval	فترة تنبؤ
Principal Components Analysis	تحليل المكونات الأساسية
Probability	احتمال
Probability distribution function	دالة توزيع الاحتمال
Proxy	مقرّب

Q

Quadratic regression Model	نموذج انحدار من الدرجة الثانية
Qualitative variable	متغير نوعي
Quantitative variable	متغير كمي
Quartile	رِبيع

R

Random component	مكون عشوائي
Random variable	متغير عشوائي
Randomness	عشوائية
Range	مدى
Rank of a matrix	رتبة مصفوفة
Reduced Model	نموذج مخفّض
Regression analysis	تحليل الانحدار
Regression function	دالة انحدار
Regression sum of squares	مجموع مربعات الانحدار
Rejection region	منطقة الرفض

Residual sum of squares	مجموع مربعات البواقي
Residuals	بواقي
Response	متغير استجابة
Ridge regression	انحدار التل

S

Sample	عينة
Sampling techniques	أساليب معاينة
Scalar	ثابت
Scatter diagram	شكل انتشار
Serial correlation	ارتباط تسلسلي
Significance level	مستوى معنوية
Simple linear correlation	ارتباط خطي بسيط
Simple linear regression	انحدار خطي بسيط
Singular matrix	مصفوفة مفردة
Specification error	خطأ توصيف
Standard deviation	انحراف معياري
Standard error	خطأ معياري
Standardized regression model	نموذج الانحدار المعياري
Standardized ridge regression	انحدار التل المعياري
Statistic	إحصاء
Statistical Analysis System (SAS)	نظام التحليل الإحصائي (ساس)
Statistical inference	استدلال إحصائي
Statistical package for social sciences (SPSS)	حزمة برامج الإحصاء للعلوم الاجتماعية
Studentized deleted residuals	بواقي ستودنت المحذوفة
Symmetrical matrix	مصفوفة متماثلة

T

T-distribution	توزيع تي
Test	اختبار
Time-series data	بيانات سلسلة زمنية
Transpose matrix	مبدلة مصفوفة
Type I error	خطأ من النوع الأول
Type II error	خطأ من النوع الثاني

U

Unbiased estimate	تقدير غير متحيز
Unbiasedness	عدم تحيز
Unique	فريد

V

Variable	متغير
Variance	تباين
Variance-covariance matrix	مصفوفة التغاير-التباين
Variance inflation factor	عامل تضخم التباين
Variance stabilization	تثبيت التباين
Vector	متجه

W

Weighted least squares	المربعات الصغرى المرجحة
Within groups	داخل المجموعات

Z

Zero-Order correlation	معامل الارتباط من المرتبة صفر -معامل الارتباط الخطي البسيط

المراجع

المراجع

أولاً: المراجع العربية:

١. دومنيك سالفاتور (١٩٨٢) الإحصاء والاقتصاد القياسي، دار ماكجروهيل للنشر، القاهرة، جمهورية مصر العربية، ترجمة سعدية حافظ منتصر.
٢. ثوماس س. كنيير و جيمس آر. تايلور، (١٩٨٣). بحوث التسويق: مدخل تطبيقي. الجزء الأول، تعريب عبدالرحمن دعالة بيلة و عبدالفتاح السيد النعماني، دار المريخ للنشر، الرياض، المملكة العربية السعودية.
٣. عساف، عبد المعطي محمد، (١٩٩٧)، المحددات الأساسية لسياسة الأجور و المرتبات لأجهزة الخدمة المدنية في الجمهورية اليمنية، مجلة الإدارة العامة، معهد الادارة العامة، الرياض، المجلد (٣٧)، العدد الثالث.ص ص ٤٨٣-٥٤٠.
٤. عبد الرحمن محمد أبو عمة والحسيني عبد الراضي ومحمود محمد إبراهيم هندي، (١٩٩٥)، مقدمة في المعاينة الإحصائية. دار المريخ للنشر، الرياض، المملكة العربية السعودية.
٥. عبد الرزاق أمين أبو شعر (١٩٩٧)، العينات وتطبيقاتها في البحوث الاجتماعية، الإدارة العامة للبحوث، معهد الإدارة العامة، الرياض، المملكة العربية السعودية.
٦. عدلي أبوطاحون (١٩٩٨) مناهج و إجراءات البحث الاجتماعي. الجزء الأول . الإسكندرية: المكتب الجامعي.
٧. مستشفى أبها للنساء والولادة والمراكز الصحية (١٩٩٩). بيانات عن أوزان ٥٠ طفلاً غير منشورة.
٨. مؤسسة النقد العربي السعودي (١٩٩٧). التقرير السنوي لمؤسسة النقد العربي. العدد رقم (٣٣).
٩. باشيوة، حسن عبدالله (٢٠١٣م). الإحصاء وتطبيقاته على الحزمة الإحصائية SPSS. عمان، مؤسسة الوراق للنشر والتوزيع.
١٠.محمود، إيهاب عبدالسلام (٢٠١٣م). تحليل البرنامج الإحصائي SPSS. عمان دار صفاء للنشر والتوزيع.
١١.طشطوش، سليمان محمد (٢٠٠١م). أساسيات المعاينة الإحصائية. دار الشروق للنشر والتوزيع.

ثانياً: المراجع الأجنبية:

1. Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and B.F. Csaki, eds, Second International Symposium on Information Theory, pp. 267–281, Akademiai Kiado, Budapest.
2. Amemiya, T. (1980), Selection of Regressors. International Economic Review, 21,331-354.
3. American Psychological Association (2010). Publication Manual of the American Psychological Association, 6th edition. Washington, DC.
4. Anscombe, F. J. (1973), Graphs in Statistical Analysis. The American Statistician 27: 17-22.
5. Barnett, V. (1991) Sample Survey Principles and Methods, 2nd edition, Edward Arnold, London.
6. Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.
7. Birkes, D. and Dodge, Y. (1993). Alternative methods of regression. Wiley Series in Probability and Statistics.
8. Bobko, P. (2001). Correlation and Regression: Applications for Industrial Organizational Psychology and Management. Thousand Oaks, CA: Sage publications.
9. Bowerman, B. L., R. T. O'Connell and A. B. Koehler (2005) Forecasting, time series and regression: an applied approach, Thomson Brooks/Cole, Belmont CA.
10. Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations. Journal of the Royal Statistical Society, B, 26:211-252
11. Breush, T. S. and Pagan, A. R. (1979), A Simple Test for Heterscedasticity and Random Coefficient Variation. Econometrica, Vol. 47, pp. 1287-1294.
12. Brooks, C. (2002). Introductory econometrics for finance. Cambridge University Press.
13. Burghes, D. N. and Wood, D. A. (1980) Mathematical Models in The Social, Management and Life Sciences. Ellis Horwood Ltd, London.
14. Carlberg, C. (2012) Predictive Analytics: Microsoft Excel. Que Publishing.
15. Chatfield, C. (1995), Problem Solving A Statistician Guide, Chapman & Hall, London.
16. Chatterjee, S. & Hadi, A. S. (2013). Regression Analysis by Example, 5th Edition. . New York: John Wiley.
17. Chatterjee, S. and Hadi, A. S. (1988). Sensitivity Analysis in Linear Regression. New York: John Wiley.

18.Cochran, W. G. (1977), Sampling Techniques, Third Edition, New York: John Wiley & Sons, Inc.

19.Cochrane, D. and Orcutt, G. H., (1949). Application of Least Squares Regressions to Relations Containing Autocorrelated Error Terms. Journal of the American Statistical Association, vol. 44, pp.32-61

20.Cody ،R. P. and Smith, J. K . (2005). Applied Statistics and the SAS Programming Language. Fifth Edition. Englewood Cliffs, NJ: Prentice Hall SAS

21.Cohen, J.; Cohen, P.; West, S.G. and Aiken, L. S. (2003).Applied Multiple Regression-Correlation Analysis for the Behavioral Sciences, 3rd Edition. Hillsdale, NJ:LEA.

22.Cook, R. D. (1977). Detection of Influential Observations in Linear Regression. Technometrics, 19: 15-18

23.Cryer, J. D., (1986), Time Series Analysis, PSW Publishers, Boston, Massachusetts 02116

24.DeMaris, A. (2004). Regression With Social Data: Modeling Continuous and Limited Response Variables. John Wiley & Sons, Inc., Hoboken, New Jersey.

25.Dobson, Annette J. (1990) An Introduction to Generalized Linear Models. Chapman and Hall, London.

26.Draper, N. R. and Smith, H. (1998). Applied Regression Analysis. 3rd edition Wiley, New York.

27.du Toit, S. H. C. , Steyn, A. G. W. and Stumpf R. H., (1986), Graphical Exploratory Data Analysis, Springer-Verlag New York Inc

28.Durbin, J. (1960), Estimation of Parameters in Time-Series Regression Models. Journal of the Royal Statistical Society, Series B, vol. 22, pp.139-153.

29.Durbin, J. and Watson, G. S, (1951), Testing for Serial Correlation in Least Squares Regression. Biometrika 38, pp. 159-178.

30.Everitt, B. S. and Dunn, G. (2010). Applied Multivariate Data Analysis (2nd edition), Wiley.

31.Eye, A. and Schuster, C.(1998). Regression analysis for social sciences Models and Applications. San Diego, CA: Academic Press.

32.Field, A. (2013). Discovering Statistics Using SPSS (4th edition). London: Sage.

33.Fox, J. (1997). Applied Regression Analysis, Linear Models, and Related Methods. Sage Publications.

34.Fox, J. and Monette, G. (1992). Generalized Collinearity Diagnostics. Journal of the American Statistical Association, 87:pp.178-183

35.Frees, E. W. (2010). Regression modeling with actuarial and financial applications. Cambridge ; New York : Cambridge University Press.

36.Freund, R. and Littell, R. (2000). SAS System for Regression. Third Edition. SAS Institute Inc., Cary, NC. USA.

37.Freund, R. and Little, R. C. (1998). SAS System for Regression. (2nd Edition). SAS Institute Inc., Cary, NC. USA.

38.Galton, F. (1886). "Regression towards mediocrity in hereditary stature". The Journal of the Anthropological Institute of Great Britain and Ireland. Vol. 15(15): 246–263.

39.George, D., & Mallery, P. (2013). IBM Statistics 21 step by step: A simple guide and reference (13 th ed.). Boston, MA: Allyn & Bacon/Prentice Hall.

40.Glejser, H., (1969), A New Test for Heteroscedasticity. Journal of the American Statistical Association, Vol. 64, pp.316-323.

41.Goldfeld, S. M. and Quandt, R. E., (1965). Some Tests for Homoscedasticity. Journal of the American Statistical Association, 60 pp.539-547.

42.Greene, W. (2006). Econometric Analysis. Upper Saddle River ،New Jersey :Prentice-Hall,. Inc.,

43.Gujarati, D. N. (1988). Basic Econometrics. (2nd edition), McGraw-Hill Book Company, New York.

44.Gujarati, D.N. and Porter, D.C. (2009) Basic Econometrics. 5th edition, New York: McGraw-Hill.

45.Gupta, D. K. (2010). Analyzing Public Policy: Concepts, Tools, and Techniques 2nd Edition. SAGE Publications

46.Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. (1975). Ridge Regression: some simulations. Communications in Statistics - Theory and Methods 4 105–123.

47.Hoerl, A.E. and Kennard, R.W. (1970a) .Ridge regression; biased estimation for non orthogonal problems. Techonometrics 12, 55-67.

48.Hoerl, A.E. and Kennard, R.W. (1970b). Ridge regression: an application to non orthogonal problems. Technometrics, 12, 69-82.

49.Hoerl, A.E. and Kennard, R.W. (1976) "Ridge regression: iterative estimation of the biasing parameter", Commun. Statist. Theor. Meth. 5(1), 77-88.

50.Howitt, Dennis (1997) Guide to Computing Statistics with SPSS for Windows. Prentice-Hall

51.Huber, P. J. (1981) Robust Statistics. New York: John Wiley and Sons.

52.Intriligator, M. D. (1978). Econometric Models, Techniques and Applications. North-Holland Publishing Company, Amsterdam. Oxford.

53.Ismail, M. (2012). Performance of Data Envelopment and Stochastic Frontier Models: In the Presence of Misspecification, Multicollinearity, and Outliers. LAP Lambert Academic Publishing.

54.Jarque, C. and Bera, A., (1987). A test for normality of observations and regression residuals. International Statistical Review 55, 163-172.

55.Jarque, C. and Bera, A., (1980). Efficient tests for normality homoscedasticity and serial independence of regression residuals. Econometric Letters 6, 255-259.

56.Johnston, J. and DiNardo, J. (1997). Econometric Methods. 4th edition. McGraw-Hill

57.Johnston, J. (1984). Econometric Methods. 3rd edition, McGraw-Hill Book Company, New York.

58.Kim, K., and Timm, N. (2007). Univariate and multivariate general linear models: Theory and applications with SAS. Chapman & Hall.

59.Kleinbaum, D. G., Kupper, L. L. & Muller, K. E. (1988). Applied Regression Analysis and other Multivariable Methods. PWS-Kent Publishing Company, Boston, second edition.

60.Kmenta, Jan (1971). Elements of Econometrics. The Macmillan Co., New York.

61.Kuiper, S., & Clippinger, D. A. (2013). Contemporary business reports (5th ed.). Mason, OH: South-Western, Cengage Learning.

62.Lewis-Beck, M. S. (editor) (1993). Regression Analysis. Sage University Paper series on Quantitative Application in Social Sciences, 07-022. Beverly Hills, CA: Sage.

63.Littell, R. C. , Stroup, W.W., and Freund, R. J. (2002). SAS for Linear Models, Fourth Edition. Cary, NC: SAS Institute Inc.SAS Institute, Inc., Cary, NC.

64.Lomax, R.G. and Hahs-Vaughn, D.L. (2012). Statistical Concepts: A Second Course. (4th Edition). Routledge: New York.

65.Maddala, G.S. (2001). Introduction to Econometrics. 3rd edition. John Wiley & Sons.

66.Mallows, C. L. (1973). Some Comments on Cp. Technometrics, 15:pp. 661-676.

67.Mallows, C.L. (1975) On some topics in robustness, Technical report, Bell Telephone Laboratories, Murray Hill, NJ.

68.Mardia, K. V. , Kent, J. T. and Bibby J. M. (1979). Multivariate Analysis. Academic Press, London.

69.Mendenhall, William and Terry Sincich (1996). A Second Course in Statistics: Regression Analysis (5TH edition). Upper Saddle River, NJ: Prentice Hall.

70.Montgomery, D. C. Peck, E. A. and Vining, G.G. (2001). Introduction to Linear Regression Analysis. 3rd Edition.Wiley Series in Probability and Statistics.

71.Montgomery, D.C., Peck, E.A., Vining, G.G. (2012). Introduction to Linear Regression Analysis (5th Edition). Wiley Series in Probability and Statistics.

72.Neter, J., Wasserman, W. and Kutner, M. H. (1990). Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. (3rd edition). Irwin, Homewood, IL 60430, Boston, MA 02116.

73.Norusis, Marija (1995). SPSS 6.1 Guide to Data Analysis. Upper Saddle, NJ: Prentice Hall.

74.Osborn, C.E. (2006). Statistical Applications for Health Information Management. Gaithersburg, MD: Aspen Publications.

75.Park, R. E., (1966), Estimation with Heteroscedastic Error Terms. Econometrica, Vol. 34, No. 4 p. 888.

76.Paulson D. S. (2007). Handbook of regression and modeling: Applications for the clinical and pharmaceutical industries. Boca Raton: Chapman & Hall/CRC. Taylor & Francis Group.

77.Puri, B. K. (1997). Statistics in Practice: Illustrated Guide to SPSS. Arnold,

78.Raykov, T., & Marcoulides, G. A. (2008). An Introduction to Applied Multivariate Analysis. New York: Taylor & Francis.

79.Ryan, T.A., Joiner, B.L., (1976). Normal probability plots and tests for normality. Minitab Statistical Software: Technical Reports. The Pennsylvania State University, State College, PA.

80.SAS Institute Inc. (1995). SAS ® User’s Guide: Basic. Version 6 ed. SAS Institute Inc., Cary, NC. USA.

81.SAS Institute Inc. (1995). SAS ® User’s Guide: Statistics. Version 6 ed. SAS Institute Inc., Cary, NC. USA.

82.Sawa, T. (1978), "Information Criteria for Discriminating among Alternative Regression Models,"Econometrica, 46, 1273–1282.

83.Schwarz, G. (1978), "Estimating the Dimension of a Model," Annals of Statistics, 6, 461–464.

84.Smith, L. F., Gratz, Z., Bousquet, S. G. (2009). The Art and Practice of Statistics. Wadsworth CENGAGE Learning.

85.Stevens, J. P. (2012). Applied Multivariate Statistics for the Social Sciences. 5th Edition Taylor & Francis.

86. Studenmund A.H. (2010). Using Econometrics: A Practical Guide. 6th Edition, Addison-Wesley Series in Economics.
87. Tabachnick, B. G. and Fidell, L. S. (2007). Using Multivariate Statistics. Pearson Education; 5th edition.
88. Theil, H. and Nagar, A. L. (1961). Testing the Independence of Regression Disturbances. Journal of the American Statistical Association, vol. 56 pp.793-806.
89. Weisberg, S. (2005). Applied Linear Regression. 3rd edition, Hoboken NJ: Wiley
90. White, H.(1980). Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48: pp.817-838
91. Wooldridge, J. M. (2009). Introductory econometrics : a modern approach . 4th ed. Mason, Ohio : South-Western Cengage Learning.
92. Yan, X. and Su, X. G. (2009). Linear Regression Analysis: Theory and Computing. World Scientific Publishing Co. Pte. Ltd. 5 Toh Tuck Link, Singapore.

المؤلف في سطور

محمد عبدالرحمن إسماعيل

- من مواليد السودان

المؤهل العلمي:

- حصل على الدكتوراه في الإحصاء من جامعة السودان للعلوم والتكنولوجيا في ٢٠١٠م، ودرجة الماجستير في الإحصاء من جامعة شفيلد بالمملكة المتحدة في ١٩٩٢م.

عمله الحالي:

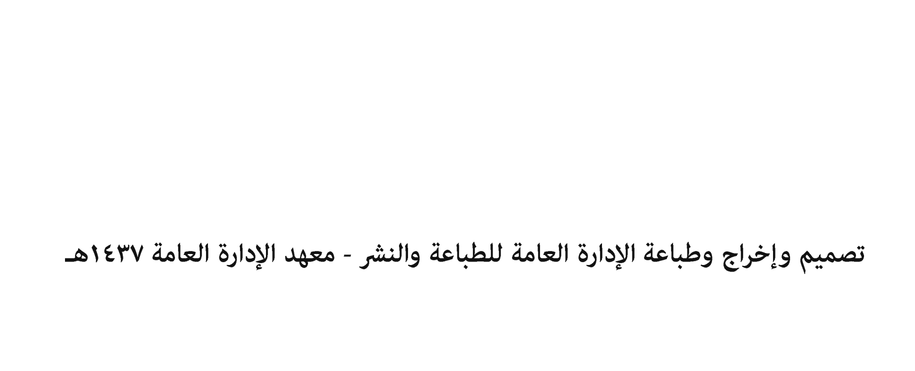
- عضو هيئة تدريب في معهد الإدارة العامة بالرياض.

الأنشطة العلمية:

Performance of Data Envelopment and Stochastic Frontier Models In the Presence of Misspecification, Multicollinearity, and Outliers. LAMBERT Academic Publishing, Germany, 2012.

- تأليف كتاب "الرقابة الإحصائية على العمليات"، ٢٠٠٦م، مركز البحوث بمعهد الإدارة العامة.
- نشر (١٢) بحثاً علمياً في دوريات علمية متنوعة و(٨) أوراق عمل قدمت في مؤتمرات، بالإضافة إلى عدد كبير من البحوث وأوراق العمل غير المنشورة قُدمت لجهات مختلفة وعدد كبير من أوراق العمل غير المنشورة.
- تدريس العديد من المواد في علم الإحصاء بكلية الاقتصاد والتنمية الريفية، وكلية التمريض بجامعة الجزيرة، وكلية الهندسة بجامعة السودان للعلوم والتكنولوجيا.

حقوق الطبع والنشر محفوظة لمعهد الإدارة العامة ولا يجوز اقتباس جزء من هذا الكتاب أو إعادة طبعه بأية صورة دون موافقة كتابية من المعهد إلا في حالات الاقتباس القصير بغرض النقد والتحليل، مع وجوب ذكر المصدر.



تصميم وإخراج وطباعة الإدارة العامة للطباعة والنشر - معهد الإدارة العامة ١٤٣٧هـ

هذا الكتاب

تعتبر نماذج الانحدار من أهم وأقوى طرق التحليل الإحصائي في البحث العلمي، والتي تهدف إلى دراسة اعتماد متغير واحد يعرف بالمتغير التابع على متغير واحد أو أكثر تعرف بالمتغيرات المفسرة أو المتغيرات المستقلة. وتستخدم نماذج الانحدار في معظم أنواع البحث العلمي – البحث التجريبي وشبه التجريبي والمُشاهد- التي غالباً ما تضم متغيرات تابعة يمكن التنبؤ بها من المتغيرات المفسرة.

يتناول هذا الكتاب موضوعات تحليل الانحدار الخطي من خلال عرض شامل وسهل ومترابط؛ بغرض تنمية مهارات النمذجة الرياضية باستخدام هذا الأسلوب. ويبدأ الكتاب بعرض بعض المفاهيم الإحصائية المهمة التي تشكل الركيزة الأساسية لموضوعات الفصول اللاحقة. وقد تناول الفصل الثاني نموذج الانحدار الخطي البسيط الذي يهدف إلى تحليل العلاقة بين متغيرين أحدهما تابع والآخر مستقل. ويعالج الفصل الثالث نموذج الانحدار الخطي المتعدد الذي يستخدم للتنبؤ بقيم المتغير التابع باستخدام متغيرين مستقلين أو أكثر. وتضمن هذا الفصل طريقة تقدير معالم النموذج، والاستدلال الإحصائي، والتنبؤ، وفحص النموذج بالإضافة إلى موضوعات أخرى ذات علاقة. وفي الفصل الرابع تمت معالجة موضوع المشاهدات الشاذة من حيث طرق كشفها وقياس أثرها وبعض طرق معالجتها. أما الفصل الخامس فقد تم تخصيصه لموضوع استخدام المتغيرات النوعية في نموذج الانحدار الخطي حيث تم التعرض إلى طرق ترميز المتغيرات الصورية وبناء نماذج انحدار تضم متغيرات نوعية وكمية ومتغيرات تفاعل بينهما. ويعالج الفصل السادس موضوع اختيار "أفضل" نموذج انحدار عندما يكون لدى الباحث مجموعة كبيرة من المتغيرات المستقلة بهدف الوصول إلى نموذج يضم عدداً قليلاً من هذه المتغيرات ويعطي أعلى درجة من الدقة في التنبؤ بالمتغير التابع. وأما الفصل السابع فيعالج أهم مشكلات عدم استيفاء اشتراطات طريقة المربعات الصغرى الاعتيادية وطرق معالجتها، فعرض الفصل أهم أربع مشكلات، هي: أخطاء توصيف نموذج الانحدار، والارتباط الخطي المتعدد، واختلاف التباين، والارتباط الذاتي. وتتميز الطبعة الثانية بإضافة فصل ثامن خُصِّص لطرق تأكيد صحة نموذج الانحدار الخطي وعرض نتائجه. فبالإضافة للمعالجة النظرية لنموذج الانحدار الخطي تم التركيز في هذا الكتاب على أمثلة وتطبيقات عملية على بيانات معظمها حقيقية باستخدام أوسع حزم برامج الإحصاء استخداماً في التحليل الإحصائي(نظام SAS وSPSS).


9660142469

تصميم وإخراج وطباعة الإدارة العامة للطباعة والنشر ـ معهد الإدارة العامة ١٤٣٧هـ